

Cloud Performance Benchmark Series

Amazon EC2 CPU Speed Benchmarks

Kalpita Sarda

Sumit Sanghrajka

Radu Sion

ver. 0.7



1. Overview

We document here an experimental evaluation of the throughput and costs of the Amazon EC2 compute infrastructure.

2. Setup

At the time of writing, Amazon EC2 provides 8 different types and On-Demand instances to work on. The Amazon results were compared with two separate local lab setups: an Intel Core 2 Duo laptop and a dual-CPU Intel Xeon CPU Server. The table below shows associated machine configurations and hourly costs.

Instance Type & Platform	RAM (GB)	CPU/ No. of EC2 compute units	Cost per hour (USD)	
			East (US)	West US, EU & APAC
Laptop – 64	4	Intel Core 2 Duo CPU T6500 @ 2.10GHz (Total Cores: 2)	NA	NA
Server – 64	8	2 x Intel Xeon CPU X5460 @ 3.15GHz (Total Cores: 8)	NA	NA
m1.small – 32	1.7	1 EC2 compute unit*	0.085	0.095
m1.large – 64	7.5	4 EC2 compute units*	0.34	0.38
m1.xlarge – 64	15	8 EC2 compute units*	0.68	0.76
m2.xlarge – 64	17.1	6.5 EC2 compute units*	0.50	0.57
m2_2xlarge – 64	34.2	13 EC2 compute units*	1.00	1.14
m2_4xlarge – 64	68.4	26 EC2 compute units*	2.00	2.28
c1_xmedium – 32	1.7	5 EC2 compute units*	0.17	0.19
c1_xlarge – 64	7	20 EC2 compute units*	0.68	0.76

*One EC2 Compute Unit provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor.

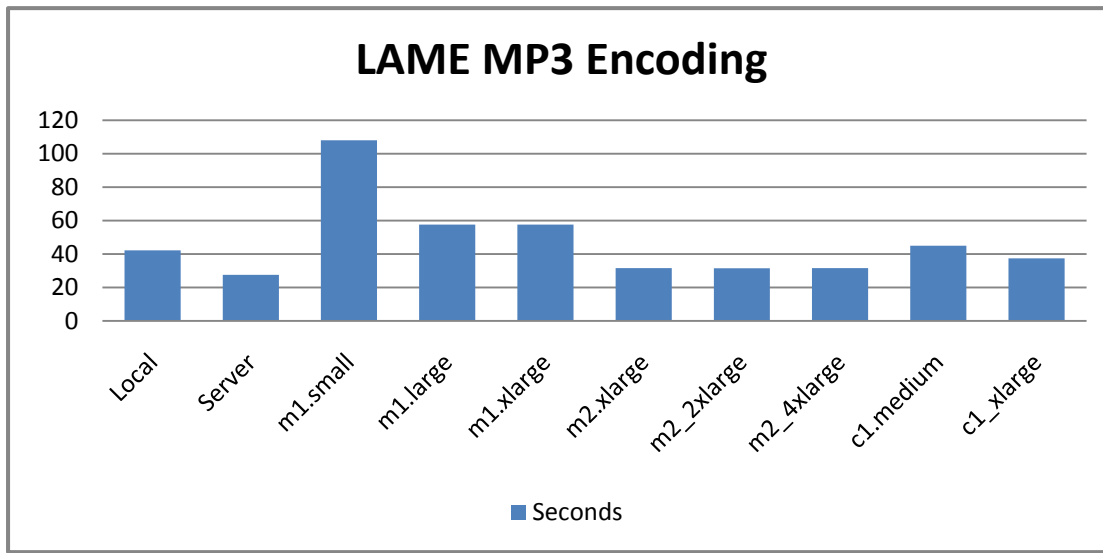
We deployed the Phoronix Test Suite to benchmark the computation throughput in a comprehensive set of scenarios and applications. Phoronix explores tens of different core applications from areas such as data compression, media encoding, cryptography, graphics rendering, scientific computation, database query processing and web serving. We also ran UBENCH to get a more thorough throughput figure at a fundamental RAM/CPU level. The allocated instances were running Ubuntu in controlled close to single-user mode. For comparison, we had run also a few Windows instances – results were virtually identical.

One of the interesting insights that can be glanced from the figures below is that the CPU performance/cost ratio is best (often by orders of magnitude) for the small instance type. This suggests that computation-rich EC2 deployments should most likely deploy multiple small instances instead of fewer large ones.

3. Phoronix Test Suite Results

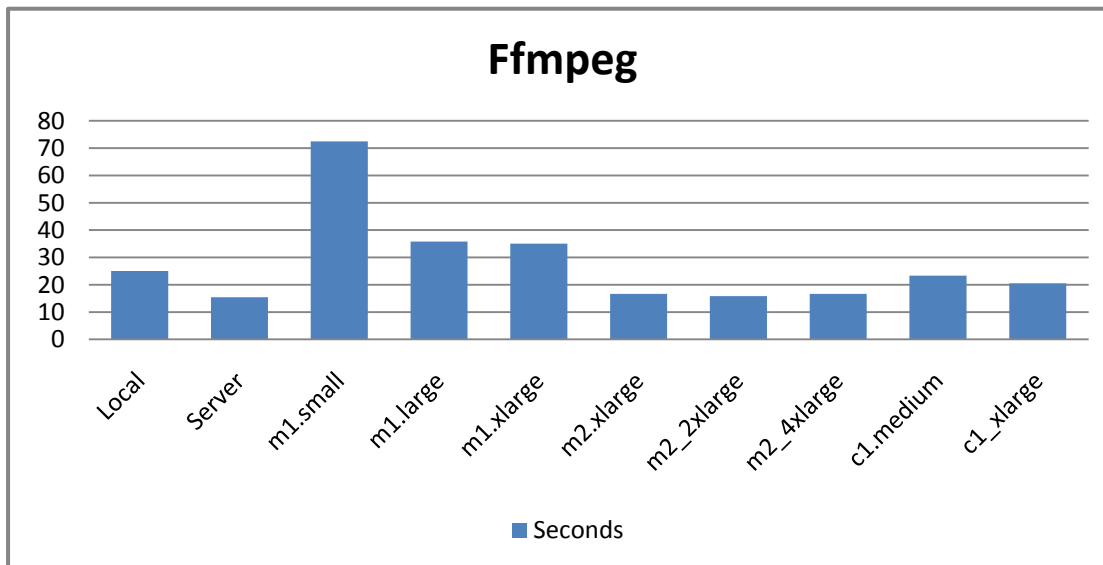
3.1. LAME MP3 Encoding

LAME is an MP3 encoder licensed under the LGPL. This test measures the time required to encode a WAV file into MP3 format. Performance is measured in seconds (lower is better).



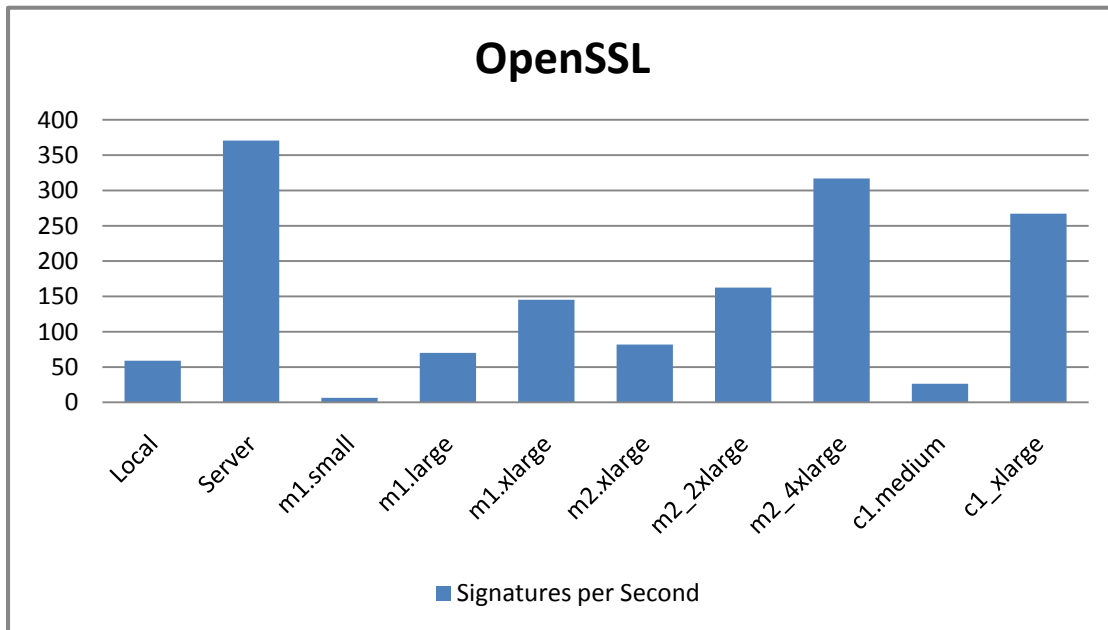
3.2. FFmpeg

This test uses FFmpeg for testing the system's audio/video encoding performance. Performance is measured in seconds (lower is better).



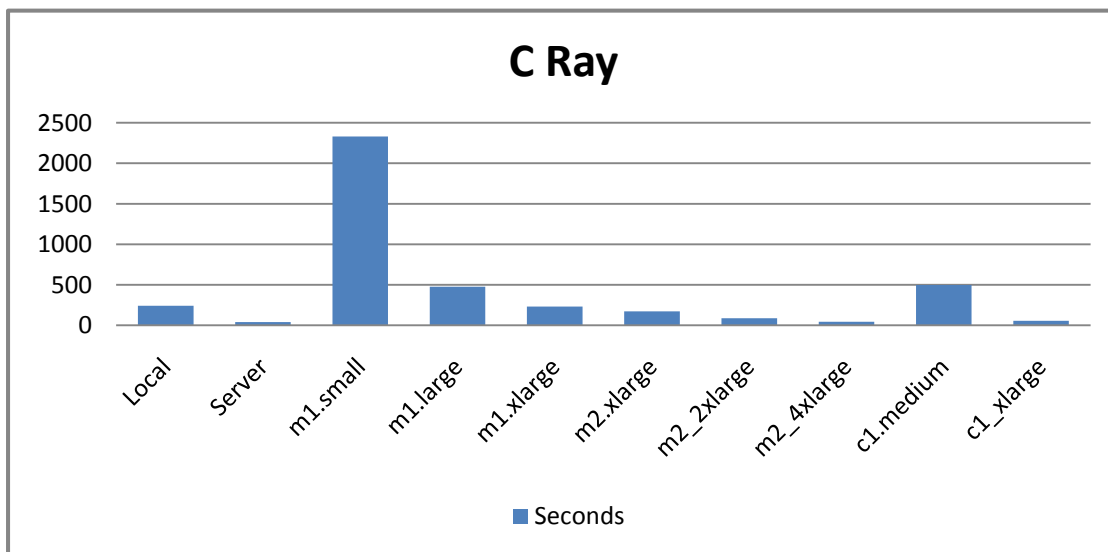
3.3. OpenSSL

OpenSSL is an open-source toolkit that implements SSL (Secure Sockets Layer) and TLS (Transport Layer Security) protocols. This test measures the RSA 4096-bit performance of OpenSSL. Performance is measured in signatures per second (higher is better).



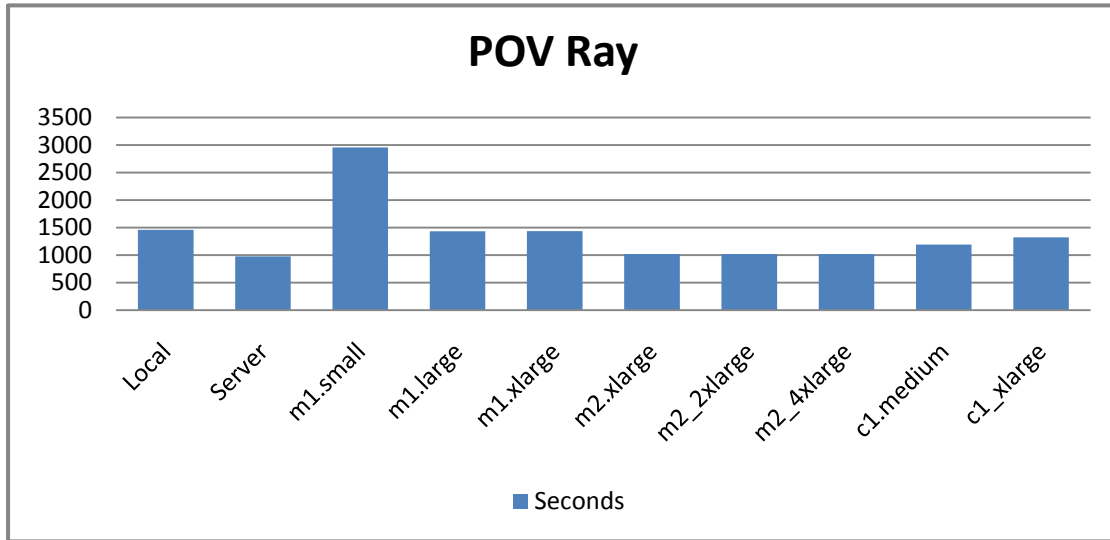
3.4. C-Ray

C-Ray, is a raytracer designed to test floating-point CPU performance. This test is multi-threaded (16 threads per core), will shoot 8 rays per pixel for anti-aliasing, and will generate a 1600 x 1200 image. Performance is measured in seconds (lower is better).



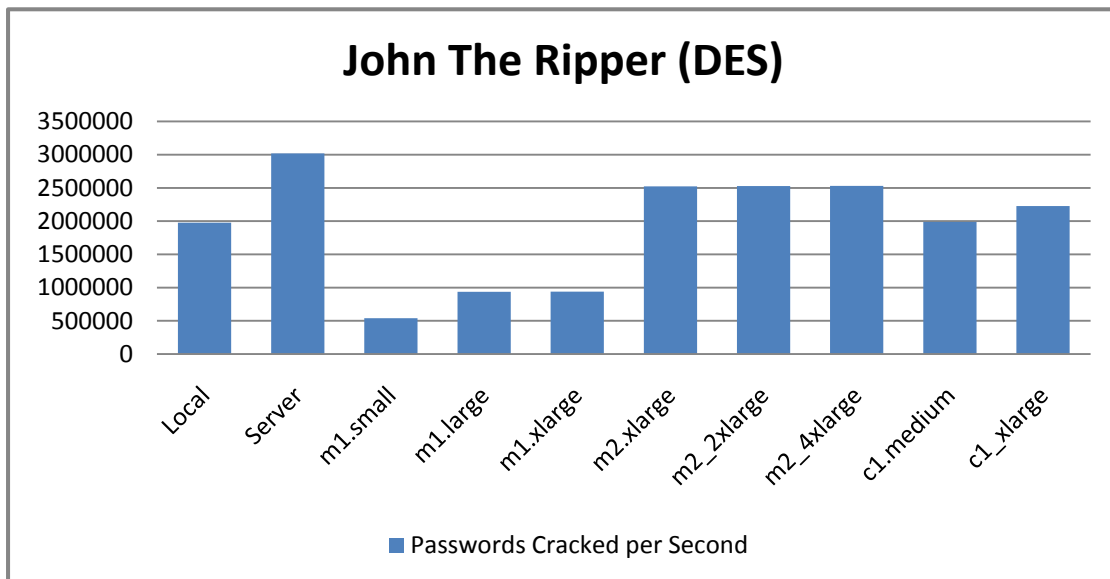
3.5. PovRay

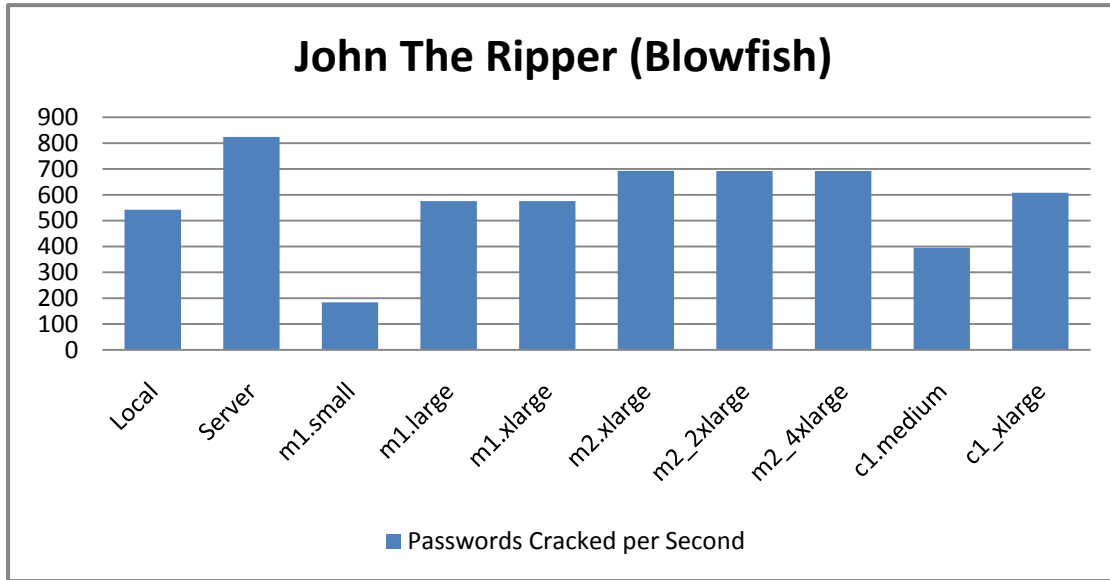
POV-Ray is the Persistence of Vision Raytracer. POV-Ray is used to create 3D graphics using ray-tracing. Performance is measured in seconds (lower is better).



3.6. John The Ripper

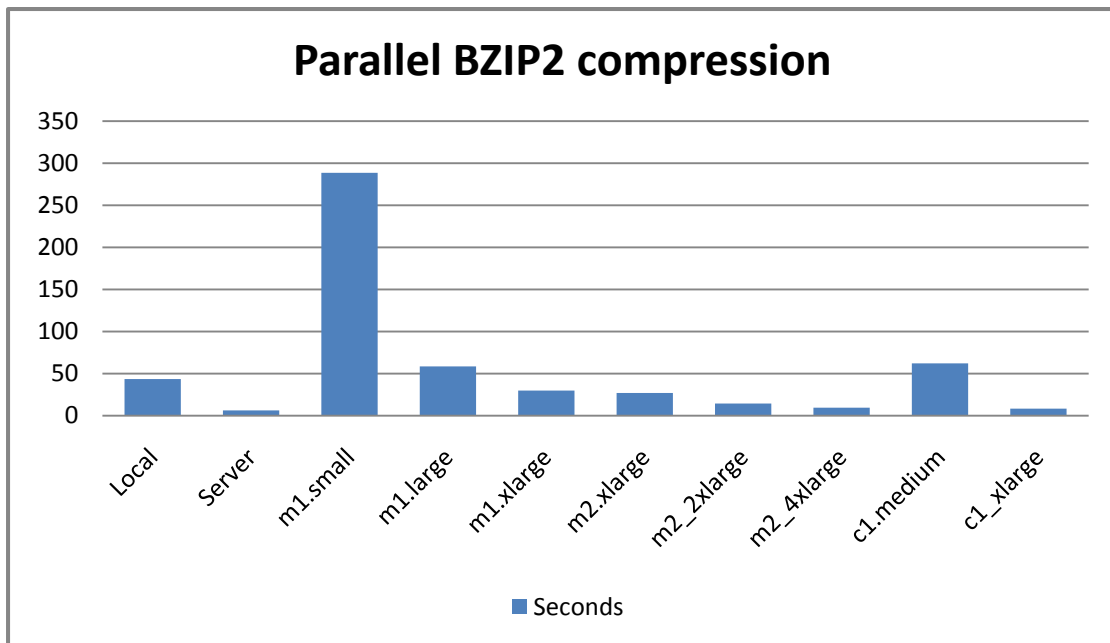
John the Ripper is a popular password cracker. Performance is measured in cracking combinations (of username and password) per second. Numbers for DES and Blowfish ciphers are shown (higher is better).





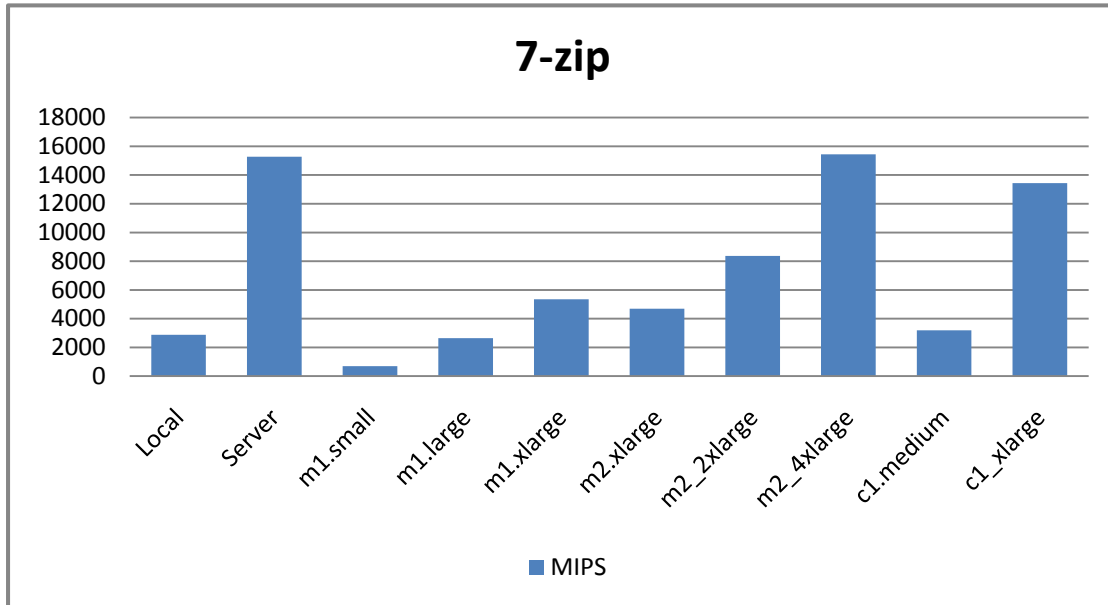
3.7. Parallel bzip2 Compression

This test measures the time needed to compress a file using BZIP2 compression deploying 16 parallel threads. A 256MB file is used for compression. Performance is measured in seconds (lower is better). Note that this also includes I/O.



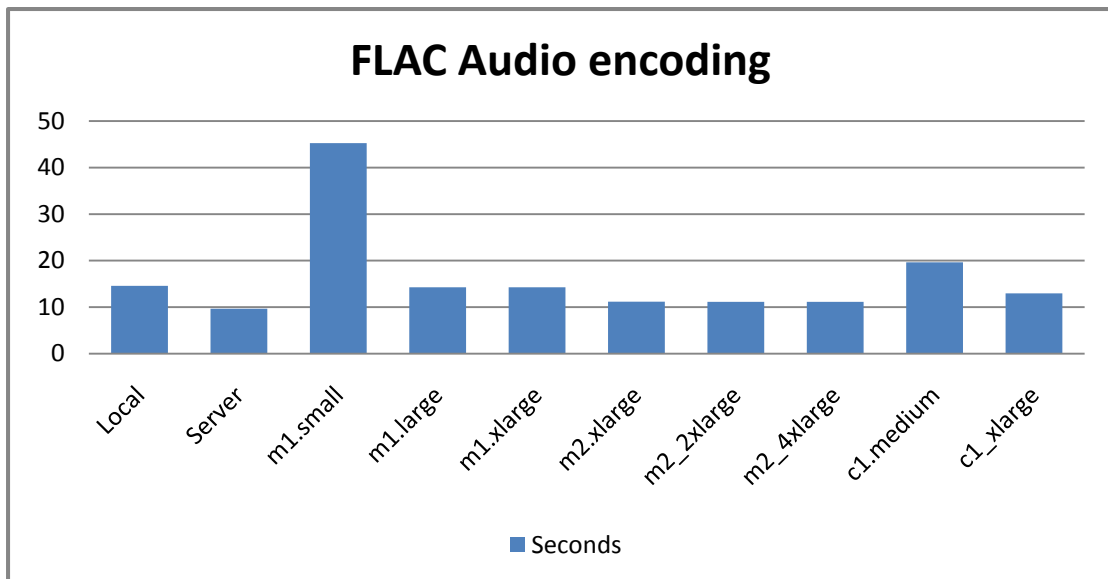
3.8. 7-Zip

This is a test of 7-Zip using p7zip with its integrated benchmark feature. Performance is measured in MIPS (higher is better).



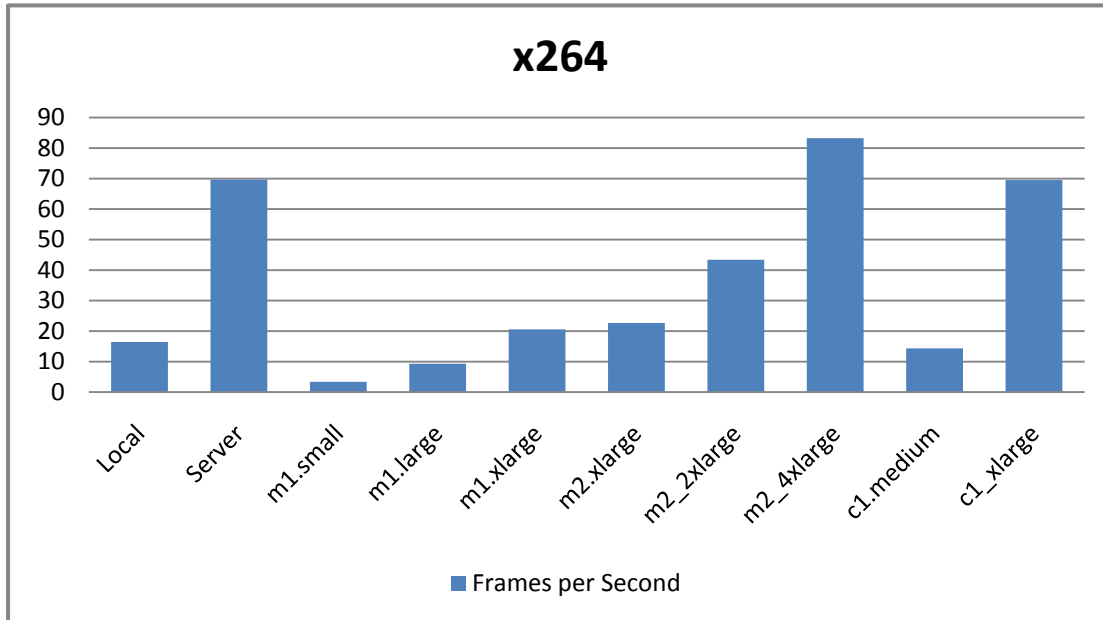
3.9. FLAC Audio Encoding

This test times how long it takes to encode a sample WAV file into a FLAC format. Performance is measured in seconds (lower is better).



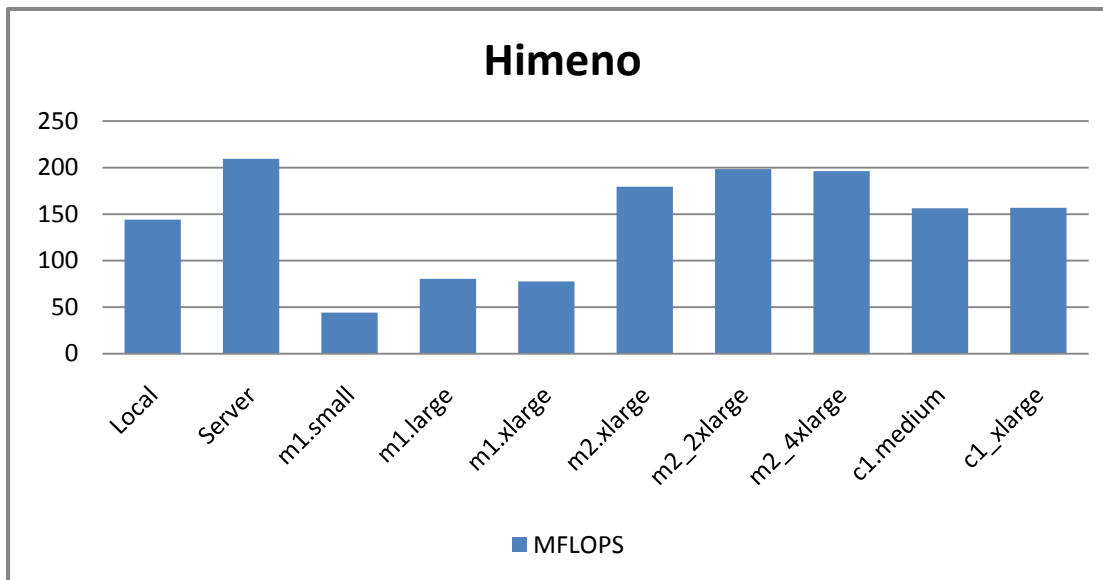
3.10. x264 encoding

This is a test of the x264 H.264/AVC encoder throughput. Performance is measured in frames per second (higher is better).



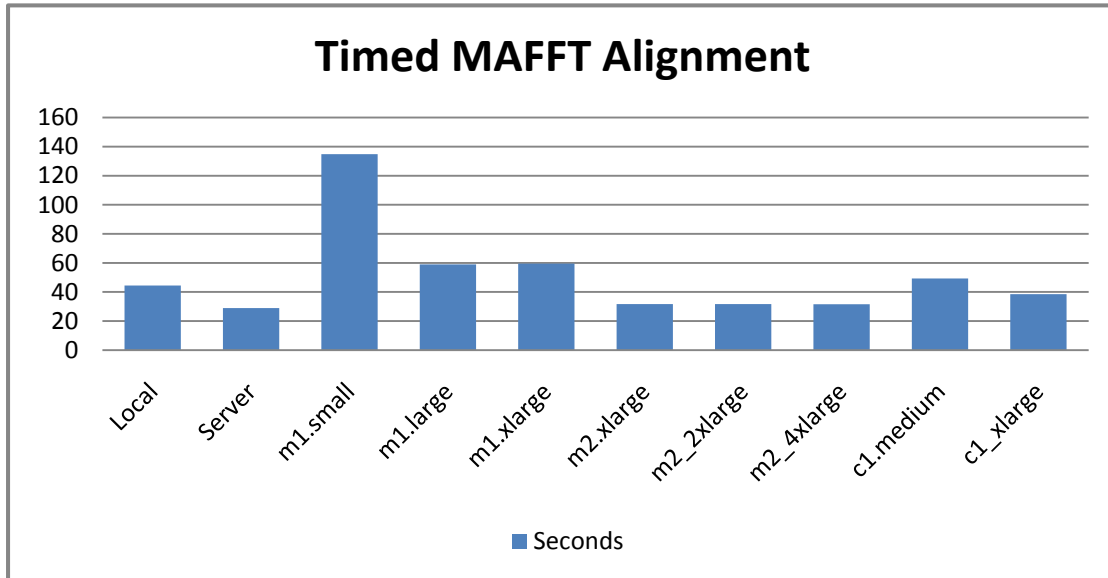
3.11. Himeno

Himeno is a linear solver of pressure Poisson using a point-Jacobi method. Performance is measured in MFLOPS (higher is better).



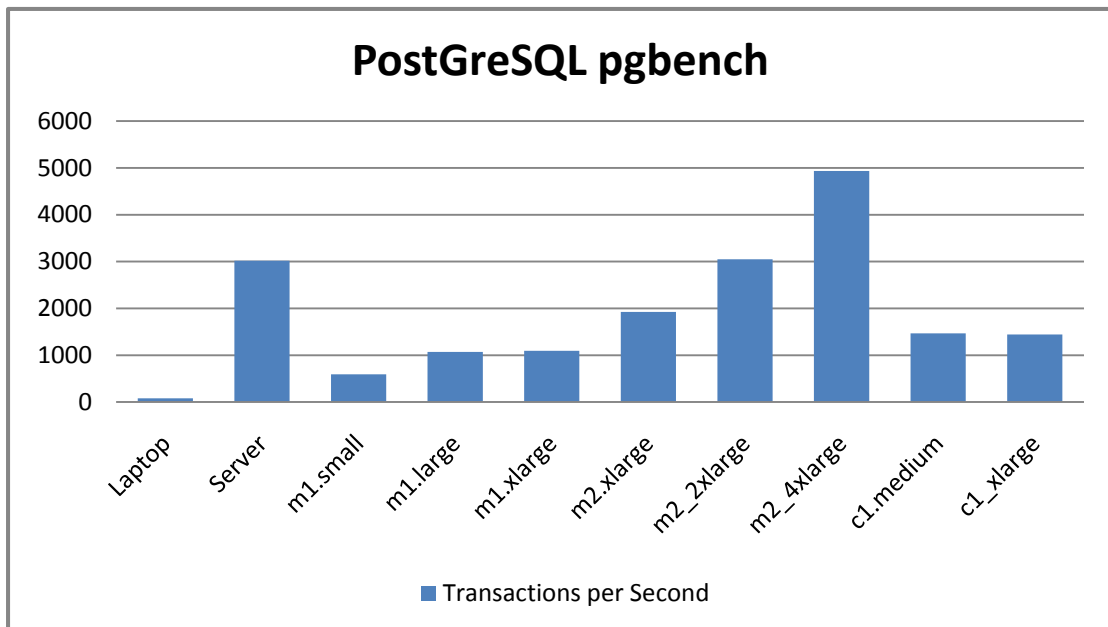
3.12. Timed MAFFT Alignment

This test performs an alignment of 100 pyruvate decarboxylase sequences using the MAFFT software. Performance is measured in seconds (lower is better).



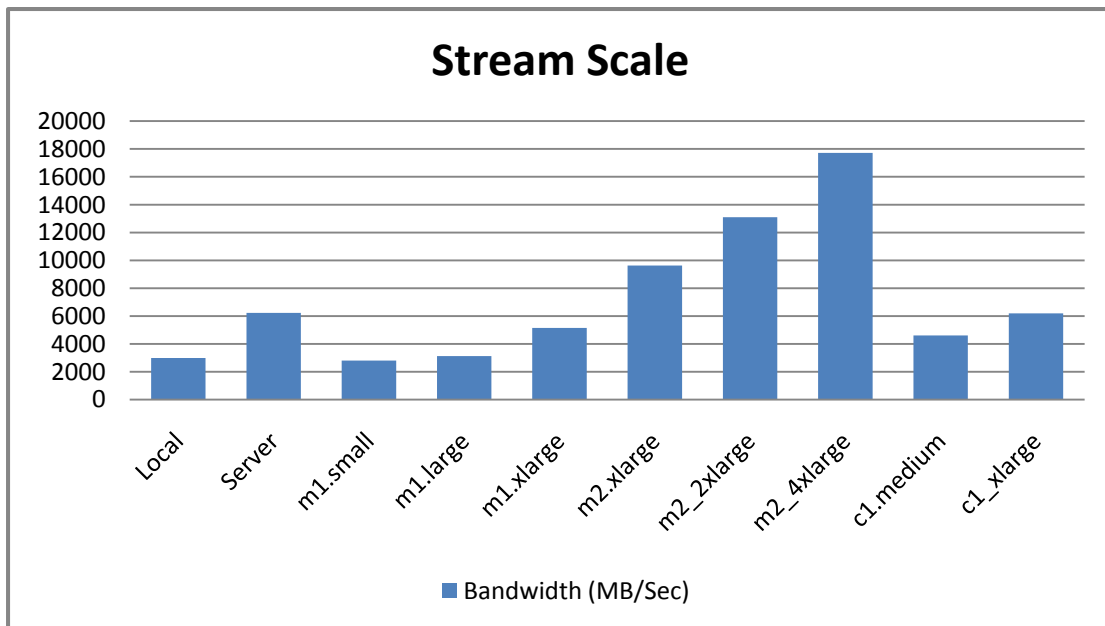
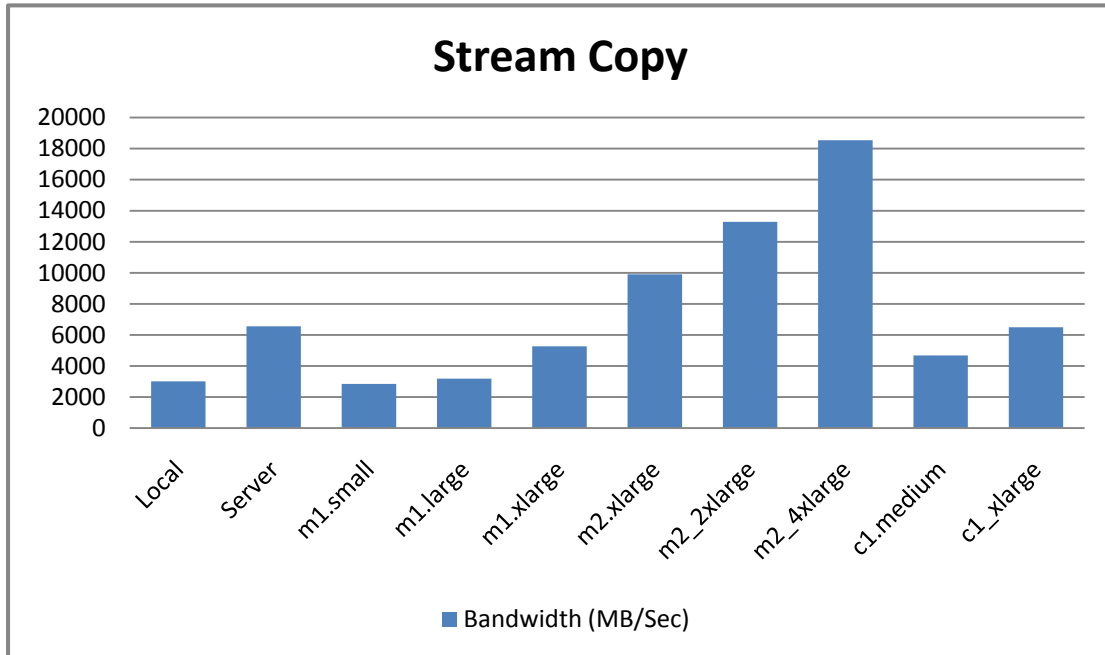
3.13. pgBench

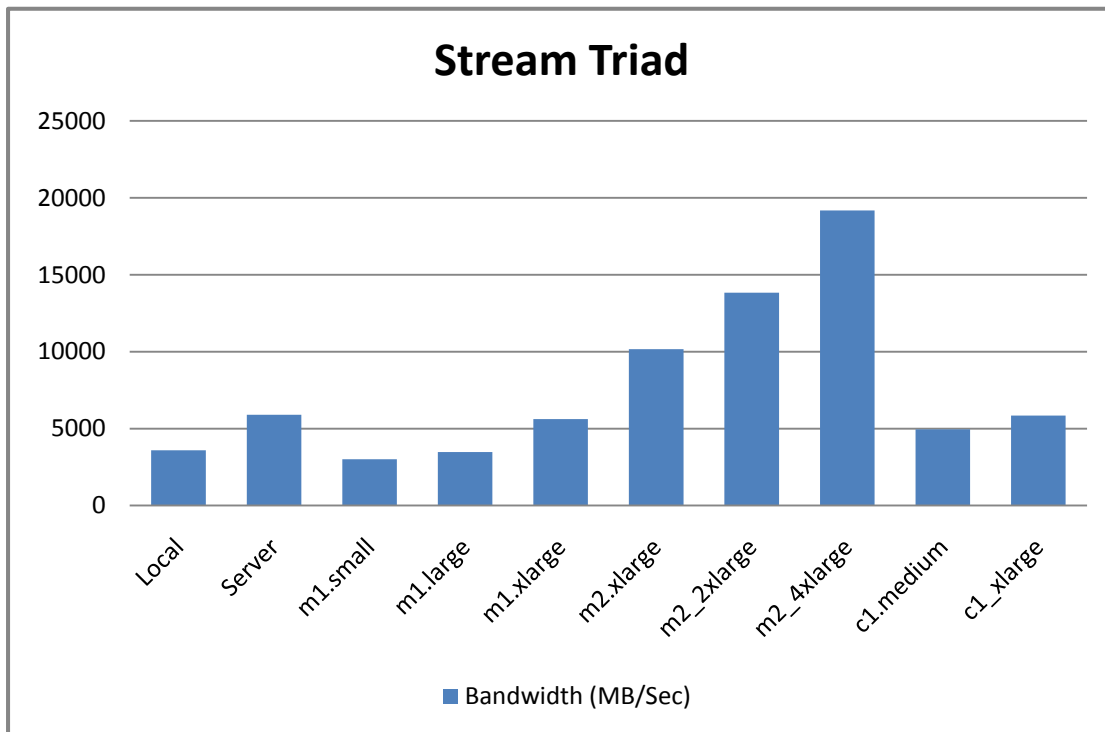
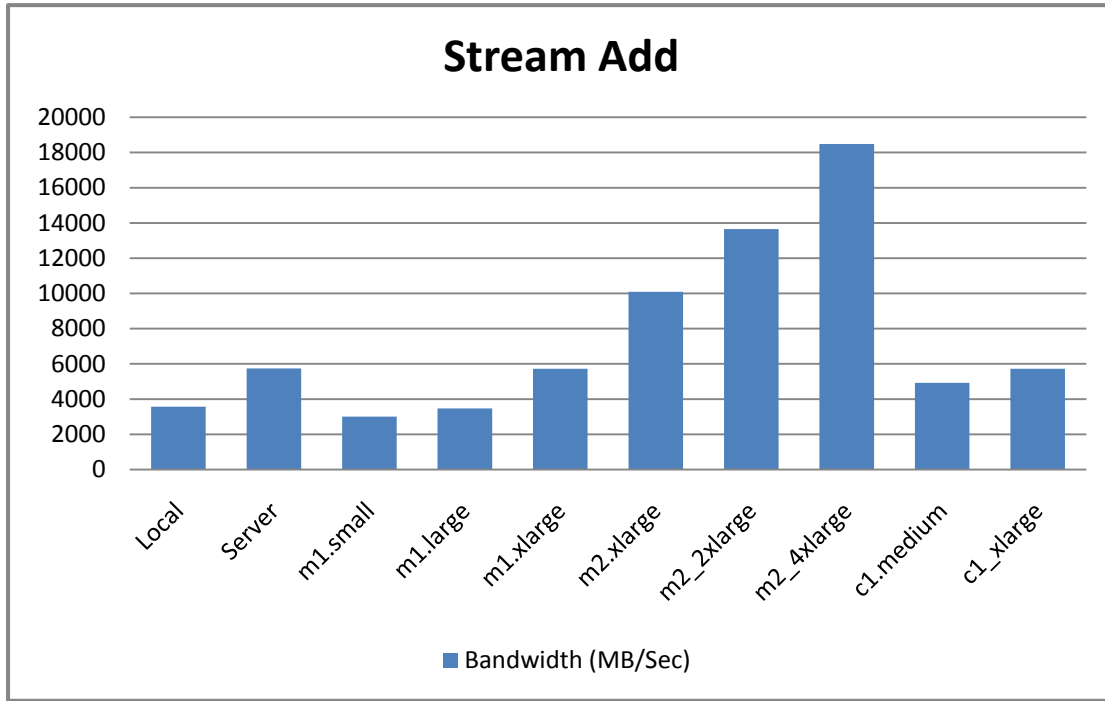
pgbench is a benchmark of PostgreSQL. It repeatedly runs a sequence of SQL commands, and then calculates the average transaction rate (transactions per second). By default, pgBench tests a scenario that is loosely based on TPC-B, involving five SELECT, UPDATE, and INSERT commands per transaction. Performance is measured in Transactions per second (higher is better). Note this includes I/O.



3.14. Stream

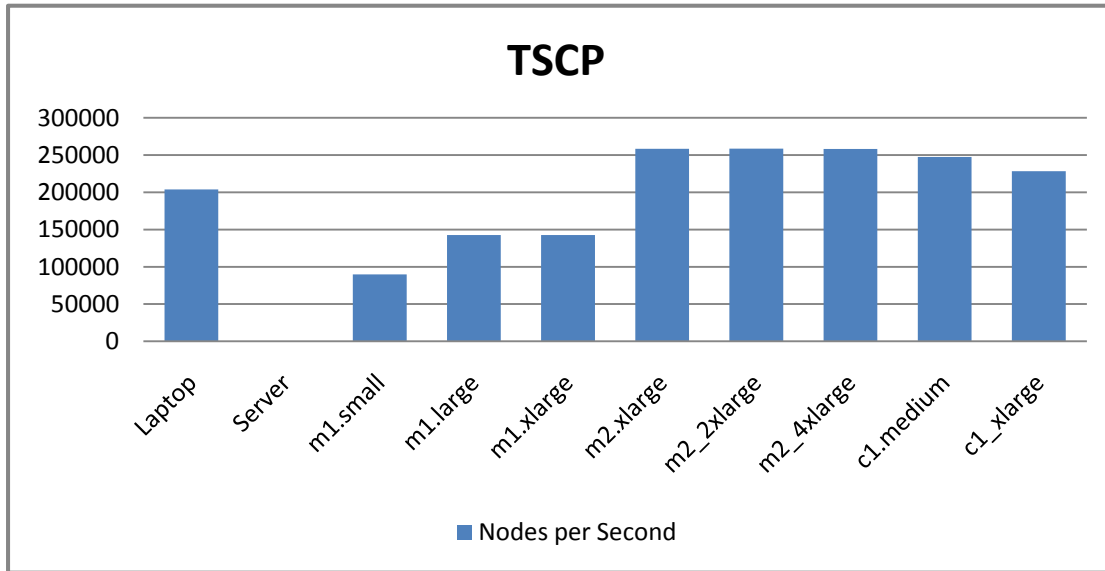
The STREAM benchmark is a synthetic benchmark that measures sustainable memory bandwidth (in MB/s) (higher is better).





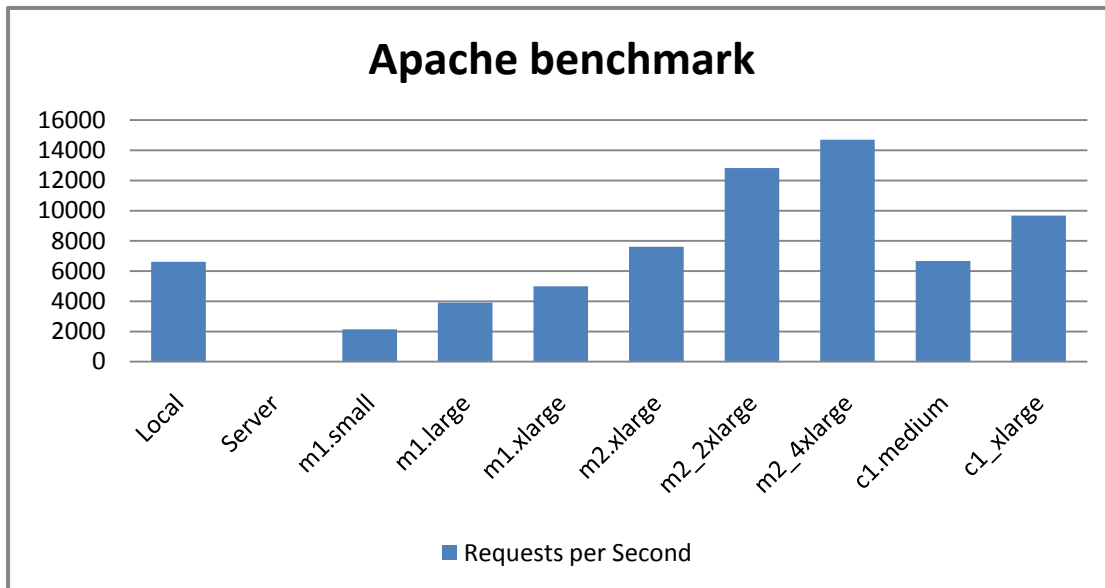
3.15. TSCP

This is a performance test of TSCP, Tom Kerrigan's Simple Chess Program, which has a built-in performance benchmark. Performance is measured in nodes searched per second (higher is better). Note: TSCP would not compile on the Xeon lab server.



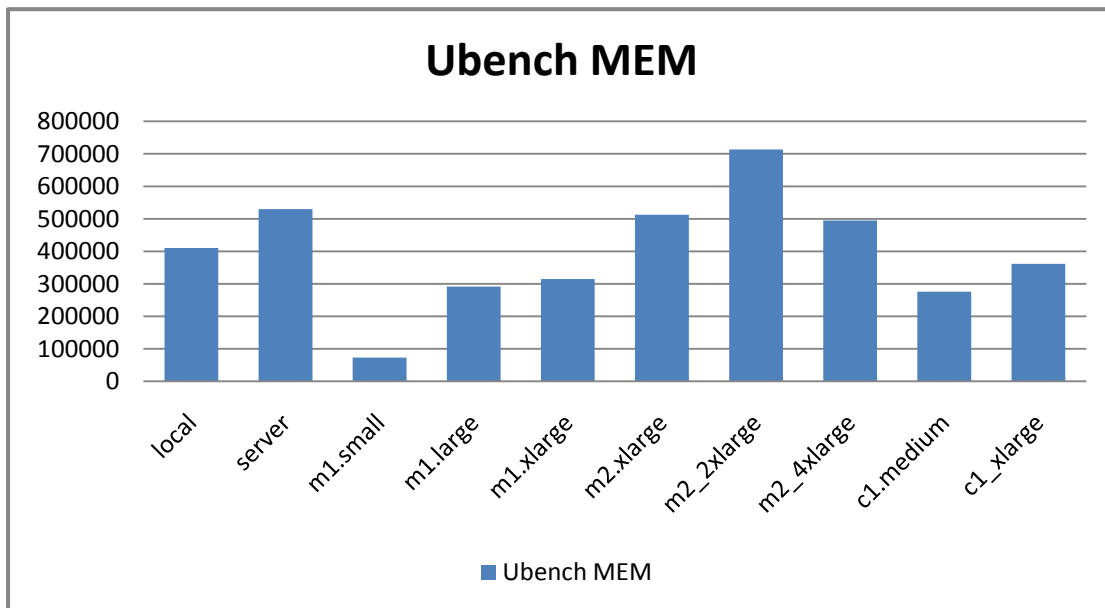
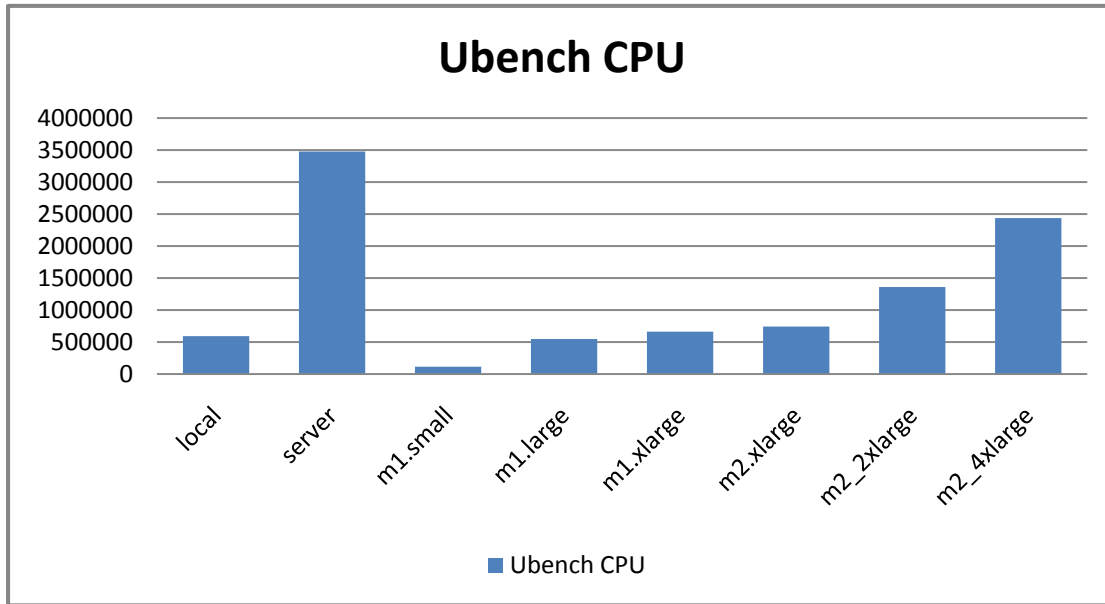
3.16. Apache Benchmark

The Apache Benchmark measures how many requests per second a given system can sustain when faced with a total of 500,000 requests, with a degree of concurrency of 100. Performance is measured in Requests per second (higher is better). Note: ab did not compile for the Xeon server.



4. UBENCH Test Suite Results

UBENCH provides a single measure of performance for machines running various unix flavors. It tests both RAM and CPU throughput. The CPU results are achieved by executing integer and floating point calculations for 3 mins concurrently using multiple threads. The ratio of floating point calculations to integer is about 1:3. The RAM results are obtained by executing allocation and copying operations for 3 mins concurrently using several threads.



5. Costs

