STAC BENCHMARK COUNCIL

STAC Update: Big Compute

Peter Nabicht
President, STAC

peter.nabicht@STACresearch.com

STAC SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-A2: Risk computation

- Non-trivial Monte Carlo calculations
  - Heston-based Greeks for multi-asset, path-dependent options with early exercise
  - Metrics: Speed, capacity, quality, efficiency
- Numerous reports
  - Some public, some in the STAC Vault
- Premium STAC members get:
  - Reports in STAC Vault
  - Detailed config info on public and private reports
  - Code from vendor implementations of the benchmarks

www.STACresearch.com/a2

STAC SECURITIES TECHNOLOGY ANALYSIS CENTER

A few points on STAC-A2 for the uninitiated

- Some tests measure **response time** for a single option of given problem size
- **Throughput** measures time to handle a portfolio of options
- **Efficiency** relates throughput to power and space
- Each response-time workload is tested 5 times, back-to-back:
  - First run is the **COLD** run
  - Subsequent 4 are **WARM** runs
- COLD relates to real-world systems that must respond to heterogeneous problem classes
  - COLD time includes building memory structures, loading kernels, etc.
- WARM relates to real-world systems configured to handle numerous requests for the same problem class

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-A2 / Dell PowerEdge XE8545 / 4 x A100 SXM4 40GB

- First STAC-A2 from Dell Technologies
- STAC-A2 Pack for CUDA (Rev G)
- Stack:
  - NVIDIA CUDA 11.6
  - Dell PowerEdge XE8545
    - 2 x AMD EPYC 7713 64-core processor @ 2.0GHz
    - 4 x NVIDIA A100 SXM4 40GiB GPU
    - 32 x 16GiB Dual Rank ECC DDR4 @ 2933Mhz
  - Red Hat Enterprise Linux 8.3

www.STACresearch.com/NVDA221007

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

Compared to all publicly reported solutions to date

- Set 3 records:
  - The highest space efficiency
    - STAC-A2.β2.HPORTFOLIO.SPACE_EFF
  - The fastest cold times in the baseline Greeks benchmark
    - STAC-A2.β2.GREEKS.TIME.COLD
  - The fastest cold times in the large Greek benchmark
    - STAC-A2.β2.GREEKS.10-100k-1260TIME.COLD

www.STACresearch.com/NVDA221007

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

Vs. a solution with 8 x GPUs*

- Had 1.2x the space efficiency
  - STAC-A2.β2.HPORTFOLIO.SPACE_EFF
- Was 2.9x the speed in the cold runs of the baseline Greeks benchmark
  - STAC-A2.β2.GREEKS.TIME.COLD
- Was 1.1x the speed in the cold runs of the Greeks benchmark
  - STAC-A2.β2.GREEKS.10-100k-1260.TIME.COLD
- Was 67% of the speed in the warm runs of the baseline Greeks benchmark
  - STAC-A2.β2.GREEKS.TIME.WARM

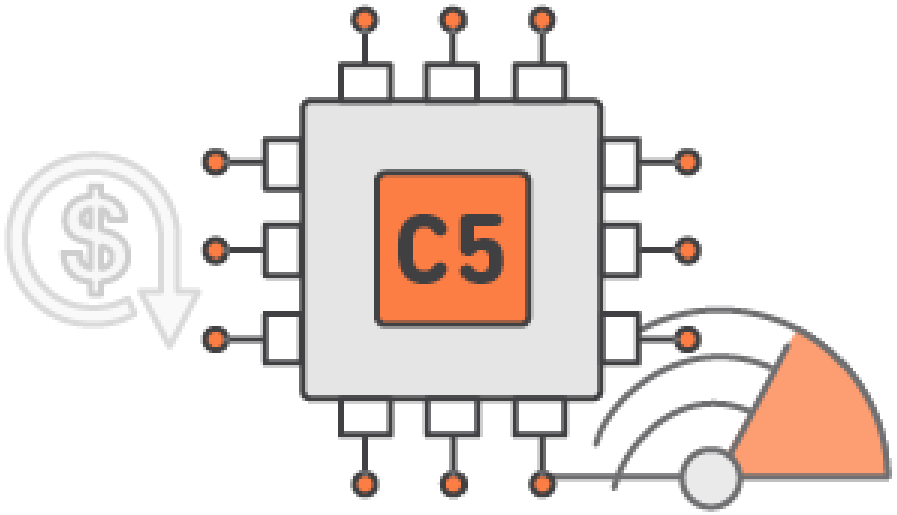
www.STACresearch.com/NVDA221007

* SUT ID NVDA210914

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-A2 / AWS c5.metal / 2 x Intel® Xeon® Platinum 8275CL

- First AWS-based solution with publicly released STAC-A2 results
- Cloud server with no hypervisor
- STAC-A2 Pack for Intel® oneAPI (Rev N)
- Stack:
  - Intel® oneAPI Base Toolkit 2022.3
  - Intel® oneAPI HPC Toolkit 2022.3
  - c5.metal Amazon Web Services Instance
    - 2 x Intel® Xeon® Platinum 8275CL (Cascade Lake) CPU @ 3.00GHz
    - 96 logical cores
    - 192 GiB of DRAM
  - Amazon Linux release 2 (Karoo)

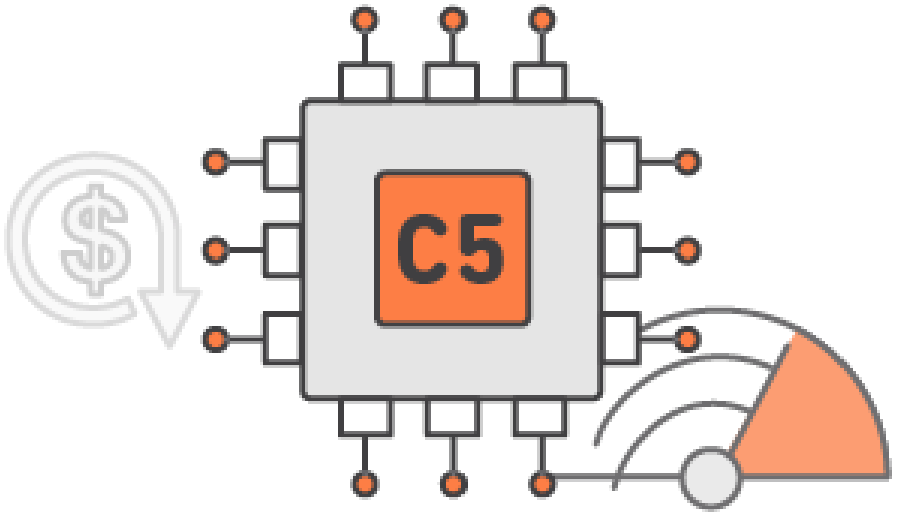


Report coming soon

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

Vs. cloud-based SUT with 10 instances and Cascade Lake CPUs*

- Completed 1.1x the options per dollar over a 1 hour burst and a 3-day period
  - STAC-A2.β2.HPORTFOLIO.PRICE_PERF .BURST and .PERIODIC
- Completed 1.2x the options per dollar (reflecting reserve instance pricing discounts) over a 1-year period
  - STAC-A2.β2.HPORTFOLIO.PRICE_PERF .CONTINUOUS



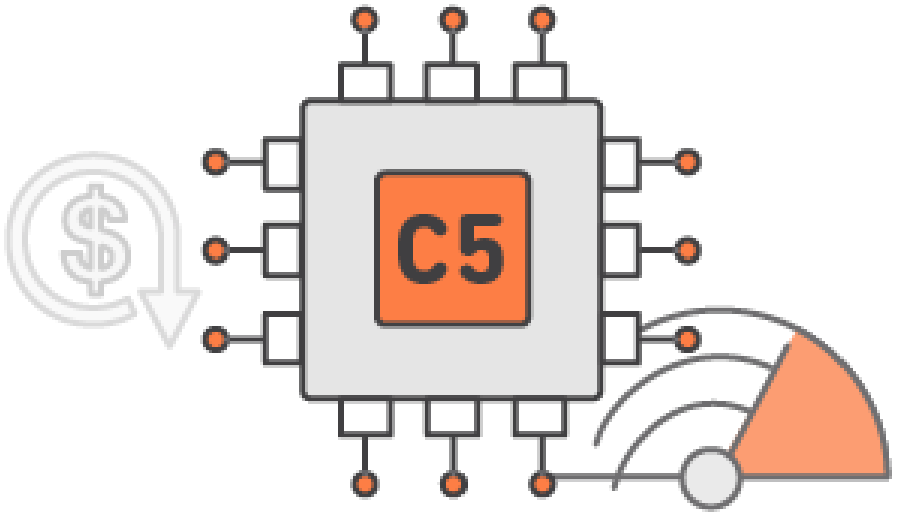
Report coming soon

* SUT ID INTC210331

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

Vs. on-prem solution with 2 x Sky Lake CPUs*

- Had 2x the throughput
  - STAC-A2.β2.HPORTFOLIO.SPEED
- In the baseline problem size
  - Was 2.4x the speed in cold runs STAC-A2.β2.GREEKS.TIME.COLD
  - Was 1.74x faster in warm runs STAC-A2.β2.GREEKS.TIME. WARM
- Was 2.3x the speed in the cold and warm runs of the large problem size
  - STAC-A2.β2.GREEKS.10-100k-1260TIME.COLD
  - STAC-A2.β2.GREEKS.10-100k-1260TIME.WARM



Report coming soon

* SUT ID INTC190401

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-A2 / AWS c6i.metal / 2 x Intel® Xeon® Platinum 8375C

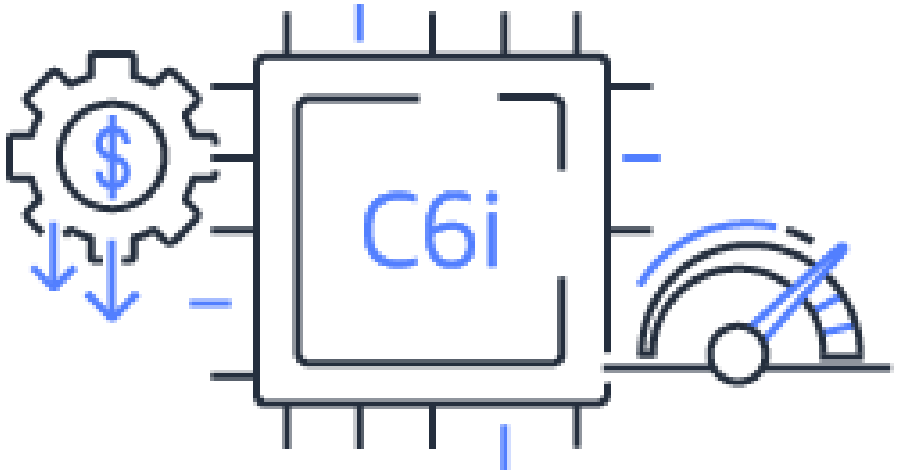
- Also, cloud server with no hypervisor but with new generation CPU (Ice Lake)
- STAC-A2 Pack for Intel® oneAPI (Rev N)
- Stack:
  - Intel® oneAPI Base Toolkit 2022.3
  - Intel® oneAPI HPC Toolkit 2022.3
  - c6i.metal Amazon Web Services Instance
    - 2 x Intel® Xeon® Platinum 8375C CPU @ 2.90GHz
    - 128 logical cores
    - 256 GiB of DRAM
  - Amazon Linux release 2 (Karoo)

Report coming soon

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

Vs. all publicly reported cloud-based solutions

- The highest options per dollar over a 1 hour burst and a 3-day period
  - STAC-A2.β2.HPORTFOLIO.PRICE_PERF.BURST
  - STAC-A2.β2.HPORTFOLIO.PRICE_PERF.PERIODIC
- The highest options per dollar (reflecting reserve instance pricing discounts) over a 1-year period
  - STAC-A2.β2.HPORTFOLIO.PRICE_PERF.CONTINUOUS

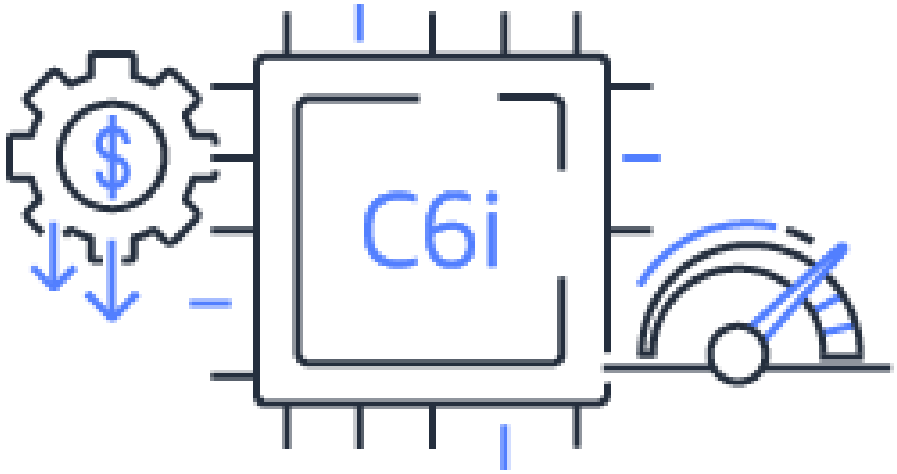


Report coming soon

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

Vs. on-prem solution with 2 x Cascade Lake CPUs*

- Had 1.5x the throughput
  - STAC-A2.β2.HPORTFOLIO.SPEED
- In the baseline problem size
  - Was 1.3x the speed in cold runs STAC-A2.β2.GREEKS.TIME.COLD
  - Was 1.4x the speed in warm runs STAC-A2.β2.GREEKS.TIME.WARM
- In the large problem size
  - Was 1.8x the speed in cold runs STAC-A2.β2.GREEKS.10-100k-1260TIME.COLD
  - Was 1.4x the speed in warm runs STAC-A2.β2.GREEKS.10-100k-1260TIME.WARM
- Handled 2 x the paths in the max paths test
  - (STAC-A2.β2.GREEKS.MAX_ASSETS)



Report coming soon

* SUT ID INTC190402

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER