

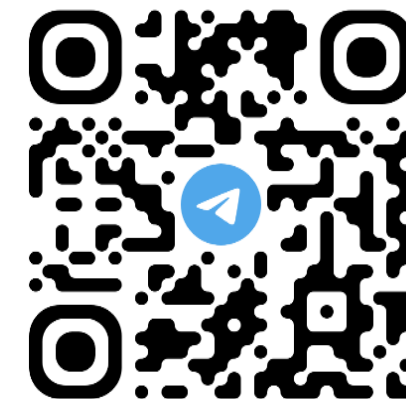


# Бенчмаркинг. Часть 1.

Лекция 2

# Содержание

- 1 Для чего нужны бенчмарки
- 2 Примеры
- 3 Бенчмаркинг в финансовой индустрии



<https://t.me/+2kGsBQZcdBQxNDAY>

# Для чего нужны бенчмарки

## Определения

- Бенчмарк – стандарт (референс), с которым происходит сравнение
- Бенчмаркинг – процесс измерения продуктов и сервисов для сравнения с продуктами и сервисами лидеров [[ссылка](#)]
  - **Технический:** для определения технических возможностей процессов и сервисов в сравнении с лидерами (пример: скорость сервиса по обработке запросов за 1 час)
  - **Конкурентный:** для сравнения процессов и сервисов с конкурентами по некоторым атрибутам и свойствам (пример: кол-во пользователей сервиса за 1 час)

# Для чего нужны бенчмарки

## Для чего?

- Для принятия решений по направлению деятельности, развитию сервисов, приобретению решений
  - **Объективные** данные по техническим характеристикам сервисов и процессов, своих или предлагаемых
  - **Объективные** данные конкурентного анализа

# Для чего нужны бенчмарки

## Проблемы разработки бенчмарков

- Нагрузки
- Сценарии использования
- Размерности задач
- Данные и их распределение
- Валидация результатов
- Таймирование
- Демонстрация возможностей программно-аппаратного стека
- Воспроизводимость
- Авторитет разработчиков, вовлеченность сообществ
- ...

# Примеры

- **HPC Challenge**, набор тестов для оценки суперкомпьютеров:
  - HPL (Highly Parallel [Linpack](#)),  $Ax=B$ , TOP500
  - STREAM – пропускная способность памяти
  - RandomAccess – скорость 64 битных обновлений случайно выбранных элементов массива
  - FFT (Fast Fourier Transform) - преобразование Фурье
  - DGEMM – матричное умножение плотных матриц
  - ...

## HPC Challenge Competition Announced

By Nicole Hemsoth  
June 2, 2005

The DARPA High Productivity Computing Systems (HPCS) Program and HPCWire are pleased to announce the first annual HPC Challenge Award Competition ([www.hpcchallenge.org](http://www.hpcchallenge.org)).

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	<b>El Capitan</b> - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	11,340,000	1,809.00	2,821.10	29,685
2	<b>Frontier</b> - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS, HPE DOE/SC/Oak Ridge National Laboratory United States	9,066,176	1,353.00	2,055.72	24,607
3	<b>Aurora</b> - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	1,980.01	38,698
4	<b>JUPITER Booster</b> - BullSequana XH3000, GH Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, RedHat Enterprise Linux, EVIDEN EuroHPC/FZJ Germany	4,801,344	1,000.00	1,226.28	15,794
5	<b>Eagle</b> - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft Azure Microsoft Azure United States	2,073,600	561.20	846.84	
6	<b>HPC6</b> - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, RHEL 8.9, HPE Eni S.p.A. Italy	3,143,520	477.90	606.97	8,461

<https://top500.org/lists/top500/2025/11/>

# Примеры

- **MLPerf**, набор тестов для сравнение моделей машинного обучения:
  - **Нагрузки:** классификация образов, обнаружение объектов, перевод
  - **Сценарии и метрики:** “single-stream”, “multi-stream”, “server”, “offline”
  - **Правила:**
    - **Разрешено:** представление данных, аппроксимации, объединение операций, использование скалярных типов
    - **Запрещено:** перетренировка, отсечение, кэширование
- **Тренировка:** высокопроизводительные вычисления
- **Прогноз:** центр обработки данных, мобильное устройство
- **Хранилище**

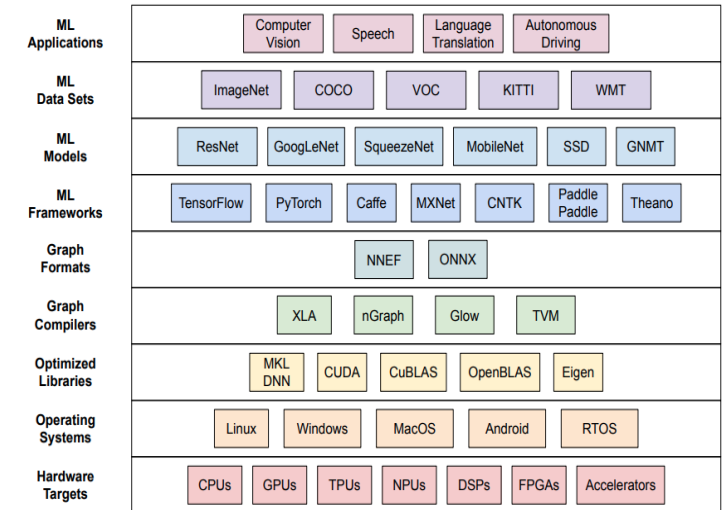


Fig. 2. The diversity of options at every level of the stack, along with the combinations across the layers, make benchmarking inference systems hard.

<https://arxiv.org/pdf/1911.02549>

MLCommons - HPC

Table - HPC (2)

MLCommons Graphs - HPC

MLCommons Stats - HPC

Division/Power

Round

Organization

System Name

Accelerator

Accelerators Per N...

Processor

Host Processors P...

MLC Model

Closed - TTT

v1.0

(All)

(All)

(All)

(All)

(All)

(All)

(All)

MLPerf Results

MLCommons data as of: 5/7/2024 7:01:22 PM

Benchmark: HPC

Division/Power: Closed - TTT

Availability: Available on-premise

Round: v1.0

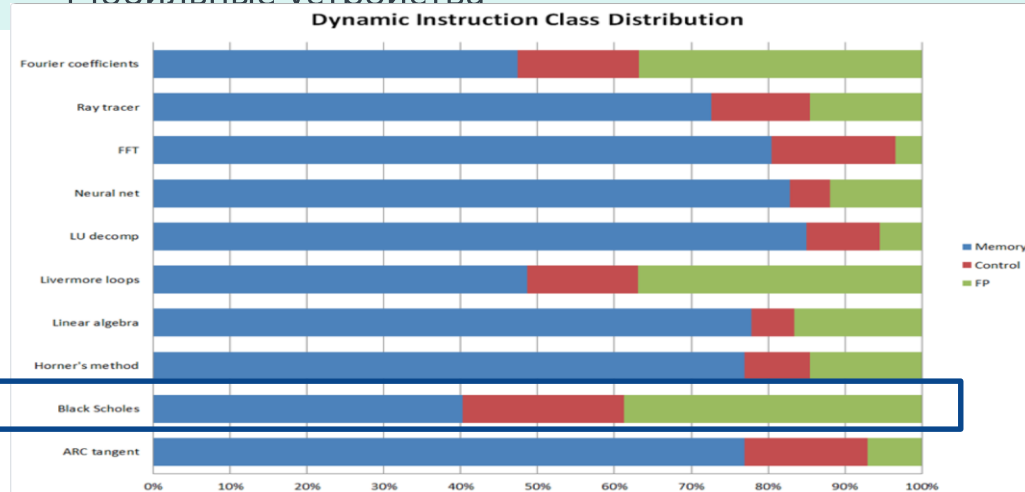
Public ID	Organization	System Name	Host Processor Model Name	Host Processors	Accelerator Model Name	Accelerators Per Node	Software
1.0-1112	HelmholtzAI	juwelsbooster_gpu_n1024_pytorc..	AMD EPYC 7402	512	NVIDIA A100-SXM4-40GB	1024	PyTorch 1.10
1.0-1113	HelmholtzAI	juwelsbooster_gpu_n2048_pytorc..	AMD EPYC 7402	1024	NVIDIA A100-SXM4-40GB	2048	PyTorch 1.10
1.0-1114	HelmholtzAI	juwelsbooster_gpu_n512_mxnnet1.9	AMD EPYC 7402	256	NVIDIA A100-SXM4-40GB	512	MXNet 1.9
1.0-1115	IBM	redpixon_4096_nv512_08	AMD EPYC 7402	512	NVIDIA A100-SXM4-40GB	512	PyTorch 1.10

https://mlcommons.org/benchmarks/training-hpc/

<https://mlcommons.org/benchmarks/training-hpc/>

# Примеры

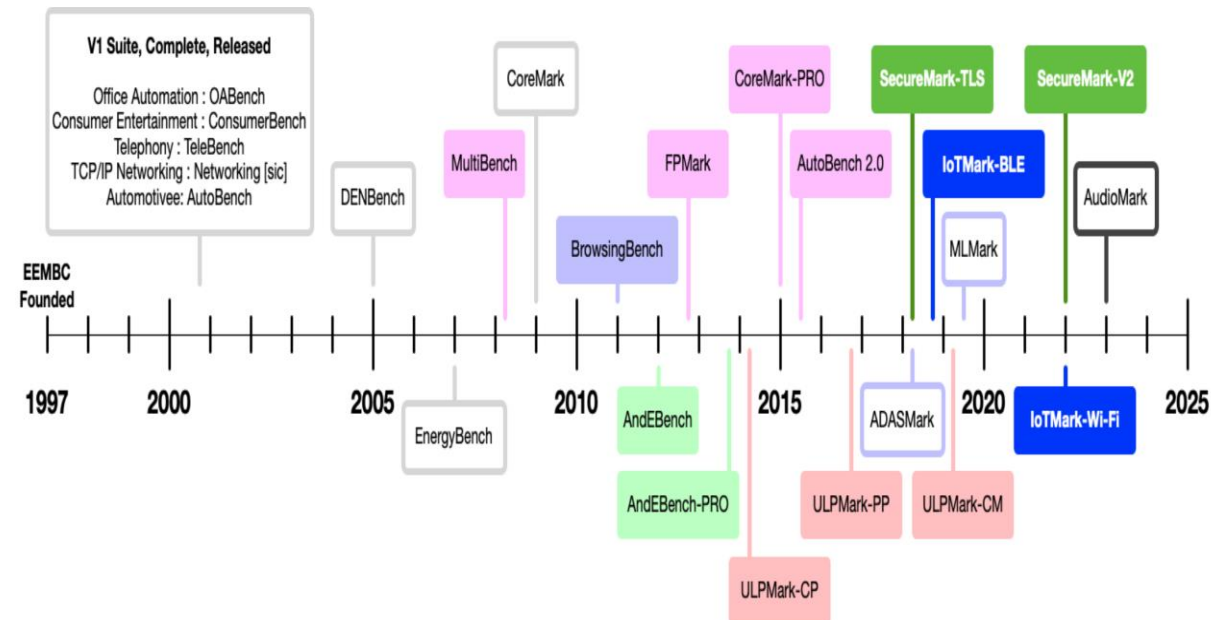
- EEMBC (EDN Embedded Microprocessor Benchmark Consortium), часть [SPEC](https://www.spec.org/) (Standard Performance Evaluation Corporation)
- [Классы](#) бенчмарков
  - Сверхнизкое энергопотребление/интернет вещей
  - Гетерогенные вычисления
  - Производительность одноядерных процессоров
  - Производительность многоядерных процессоров
  - Мобильные устройства



<https://www.eembc.org/fpmark/>

The screenshot shows the SPEC website with the URL [spec.org/benchmarks.html](https://www.spec.org/benchmarks.html). The main navigation bar includes Home, Benchmarks, Tools, Results, Contact, and Blog. The 'Benchmarks' section is expanded, listing various categories: Cloud, CPU, Embedded, Graphics/Workstations, High Performance Computing, Java Client/Server, Machine Learning, Storage, Power, Virtualization, and Results Search. The 'SPEC's Benchmarks and Tools' section lists: Current benchmarks, Tools, and Retired benchmarks. The 'Current Benchmarks' section is highlighted, showing 'Cloud'.

A brief history of EEMBC benchmark releases, note some of these benchmarks have been deprecated (e.g., EnergyBench™):



<https://www.eembc.org/products/>



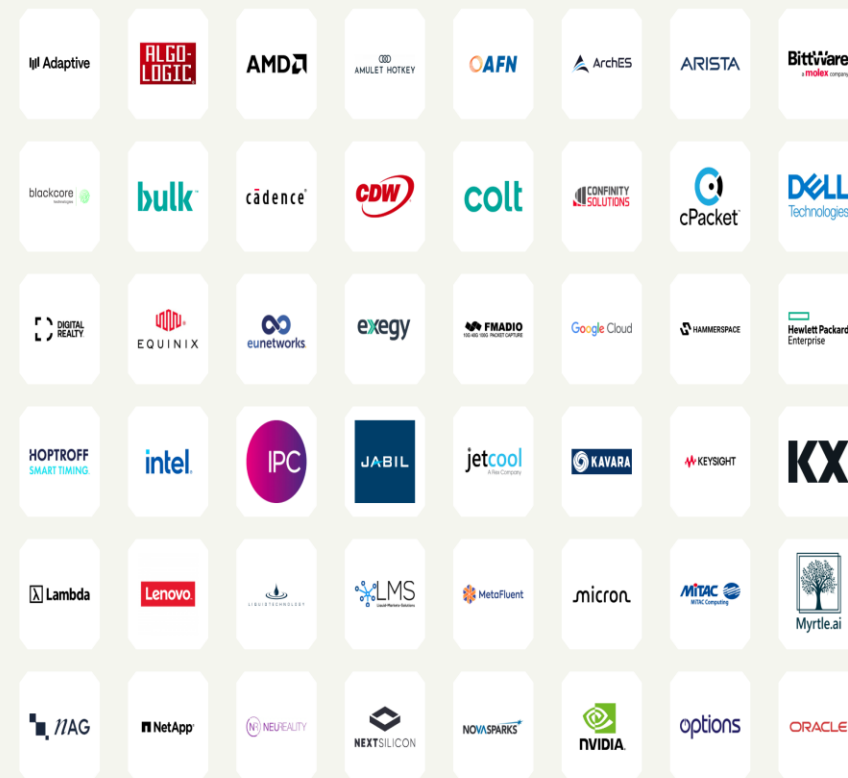
# Бенчмаркинг в финансовой индустрии

## STAC (Securities Technology Analysis Center),

<https://www.stacresearch.com/>

- 500 финансовых институтов: банки, биржи, фонды
- 70+ поставщиков технологических решений: программное обеспечение, аппаратное обеспечение, облачные решения
- Определение необходимых для финансовой индустрии технологических стандартов
- Дискуссии по техническим проблемам и их решению в финансовой индустрии

## TECHNOLOGY AND SERVICE PROVIDERS



<https://www.stacresearch.com/members>

# Бенчмаркинг в финансовой индустрии



Тема  
следующих  
занятий

## Разные классы нагрузок

- Вычисления
- Ввод-вывод для сетей
- Ввод-вывод для хранилищ данных

## Горизонтальные нагрузки для принятия решений

- Рыночные данные
- Аналитика
- Машинное обучение
- Исполнение стратегий
- Синхронизация времени

# Бенчмаркинг в финансовой индустрии

## STAC-A2

- Для тестирования технологических стеков, используемых для решения вычислительно затратных задач по управлению рисками
- Набор спецификаций
  - Математические функции (exp, log, квадратный корень)
  - Генерация независимых и коррелированных Гауссовых величин
  - Генерация цен акций
  - Вычисление «греков» (чувствительность цен производных активов к параметрам рынка)

## End-User Driven Technology Benchmarks Based on Market-Risk Workloads

Peter Lankford  
*Securities Technology Analysis Center (STAC)*

Lars Ericson  
*Catskills Research Company*

Andrey Nikolaev  
*Intel Corporation*

**Abstract-** Market risk management is a critical, resource-intensive task for financial trading firms. The industry relies heavily on innovation in technical infrastructure to increase the quality and quantity of risk management information and to

trading desks incorporate risk information into their decision making, requiring shorter turnaround times for analysis.

A firm's goals with respect to market risk analysis can depend on the extent to which it is focused on cost reduction

[Workshop on High Performance Computing in Finance at SC12 \(SuperComputing 2012\).](#)

SUT ID: INTC251024

STAC-A2

### STAC-A2 report for AWS EC2 c8i.metal-96xl Instance with Intel® Xeon® 6975P-C Processors

Type: Audited

Specs: STAC-A2 Benchmark

Stack under test:

- STAC-A2 Pack for oneAPI (Rev R)
- Intel® oneAPI Base Toolkit 2025.2
- Intel® oneAPI HPC Toolkit 2025.2
- Amazon Linux 2023
- Amazon Web Services (AWS) Elastic Compute Cloud (EC2) c8i.metal-96xl Instance
  - 2 x Intel® Xeon® 6975P-C Processors
  - 24 x 32 GiB DDR5 DIMMs @ 7200 MT/s; 768 GiB total
  - 32.2 GB Amazon Elastic Block Store

<https://docs.stacresearch.com/INTC251024>

# Домашнее задание

Для рассмотренных или иных бенчмарков продемонстрировать, как адресуются проблемы по

- репрезентативности нагрузок
- таймированию
- размерностям задач
- ...