

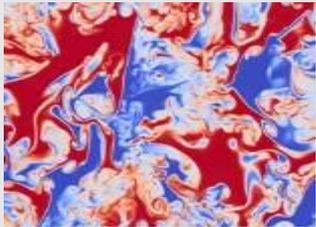
Вычисления на GPU и ИИ

Антон Джораев, adzhoraev@nvidia.com



AI SUPERCOMPUTING: НОВЫЙ ЭТАП РАЗВИТИЯ НРС

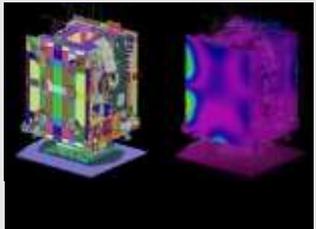
Расширение возможностей НРС с помощью искусственного интеллекта



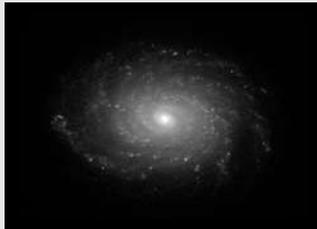
Turbulent Flow



Molecular Dynamics



Structural Analysis

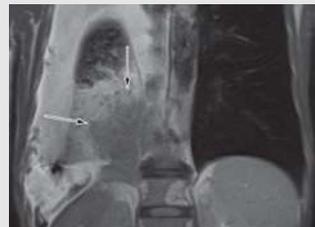


N-body Simulation

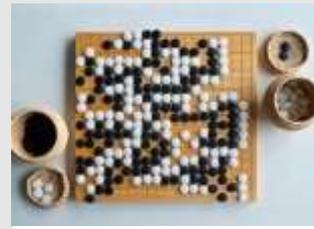
COMPUTATIONAL SCIENCE



“What’s happening?”



“Is there cancer?”

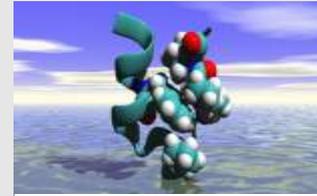


“Next move?”



“What does she mean?”

DATA SCIENCE



Drug Discovery



Clean Energy



Understanding Universe



Monitoring Climate Change

COMPUTATIONAL & DATA SCIENCE

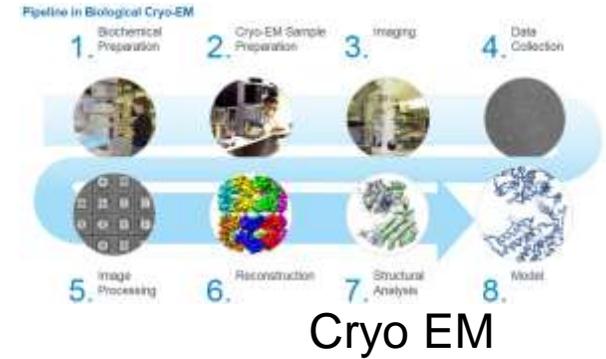
Большие эксперименты = большие данные

Обработка результатов экспериментов требует применения ИИ

How will SKA1 be better than today's best radio telescopes?

SKA1 LOW	SKA1 MID	SKA1 HIGH
RESOLUTION x12	SURVEY SPEED x135	SENSITIVITY x8
RESOLUTION x4	SURVEY SPEED x60	SENSITIVITY x5

Exabyte/Day



A GIANT
23,000 Machine weight

10X THE CORE OF THE SUN
150 million°C Plasma temperature

FUSION ENERGY
500 MW Output power

ITER TOKAMAK
ITER tokamak is an experimental machine designed to harness the energy of fusion. ITER, the world's largest tokamak, with a plasma radius (R) of 6.2 m and a plasma volume

30X Increase in power

High Luminosity LHC

10X Increase in Data Volume

How the Box Works
The Personal Genome Machine looks like a stack of consumer electronics and it uses the same old technology as a computer that can read and write DNA. Along with the PGM, you'll also have a PC, a tablet, and a smartphone to help you manage your data.

Personal Genomics

3 NVIDIA

Примеры совмещения HPC и ИИ

Deep learning дополняет и расширяет возможности HPC

Результаты вычислений используются для обучения нейросети которая дальше увеличивает точность последующих вычислений

Результаты вычислений используются для обучения нейросети которая дальше используется вместо проведения полномасштабных вычислений

HPC + AI

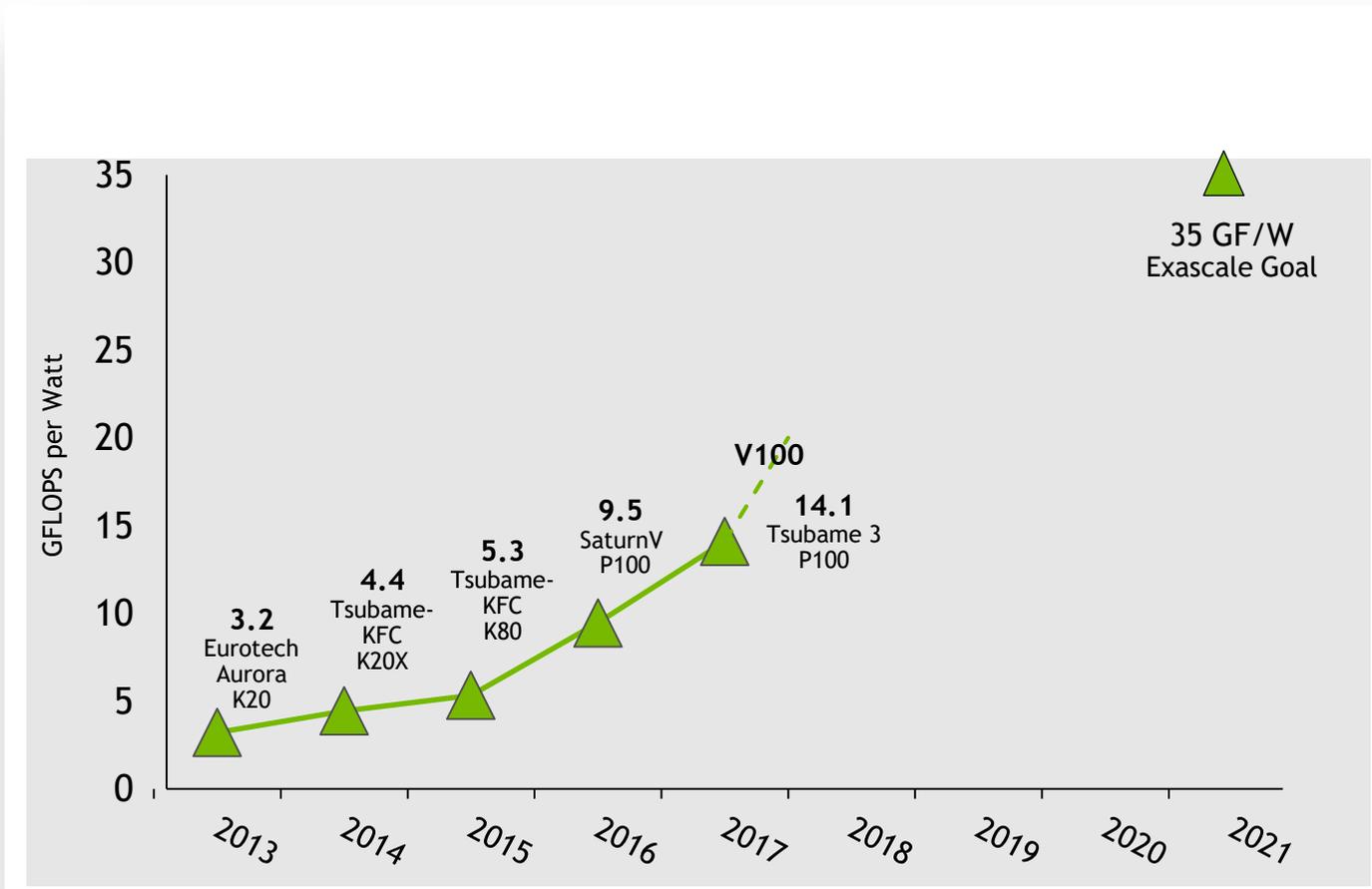
Нейросеть тренируется на результатах вычислений, а дальше используется для управления наборами исходных параметров и запусками вычислений, так чтобы минимизировать их число

Нейросеть обучается на результатах вычислений с тем чтобы на ее основе сделать систему выдающую решение в реальном времени

**Искусственный интеллект - мощное
дополнение к традиционным НРС-методам.**

ЭНЕРГОЭФФЕКТИВНОСТЬ ДЛЯ ЭКЗАФЛОПСА

GPU развиваются с опережением

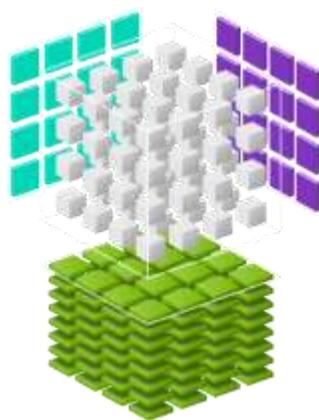


VOLTA В ОСНОВЕ КОМПЬЮТЕРА SUMMIT

новая вершина в AI Supercomputing

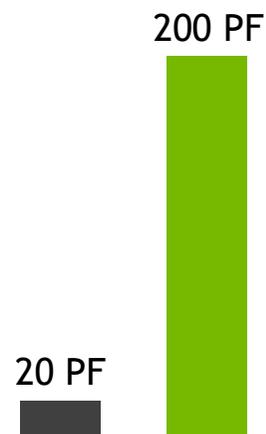


AI Exascale
уже сегодня



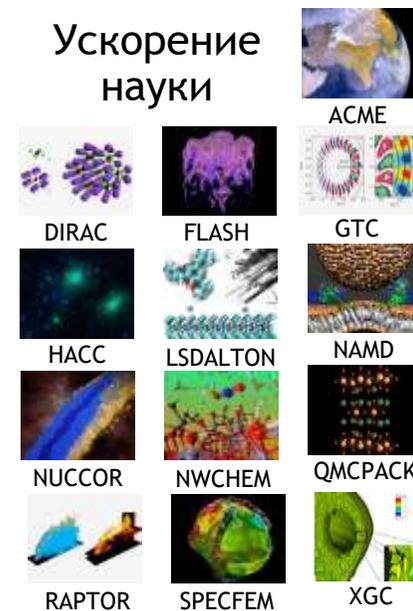
3+EFLOPS
Tensor Ops

Лидер по
производительности



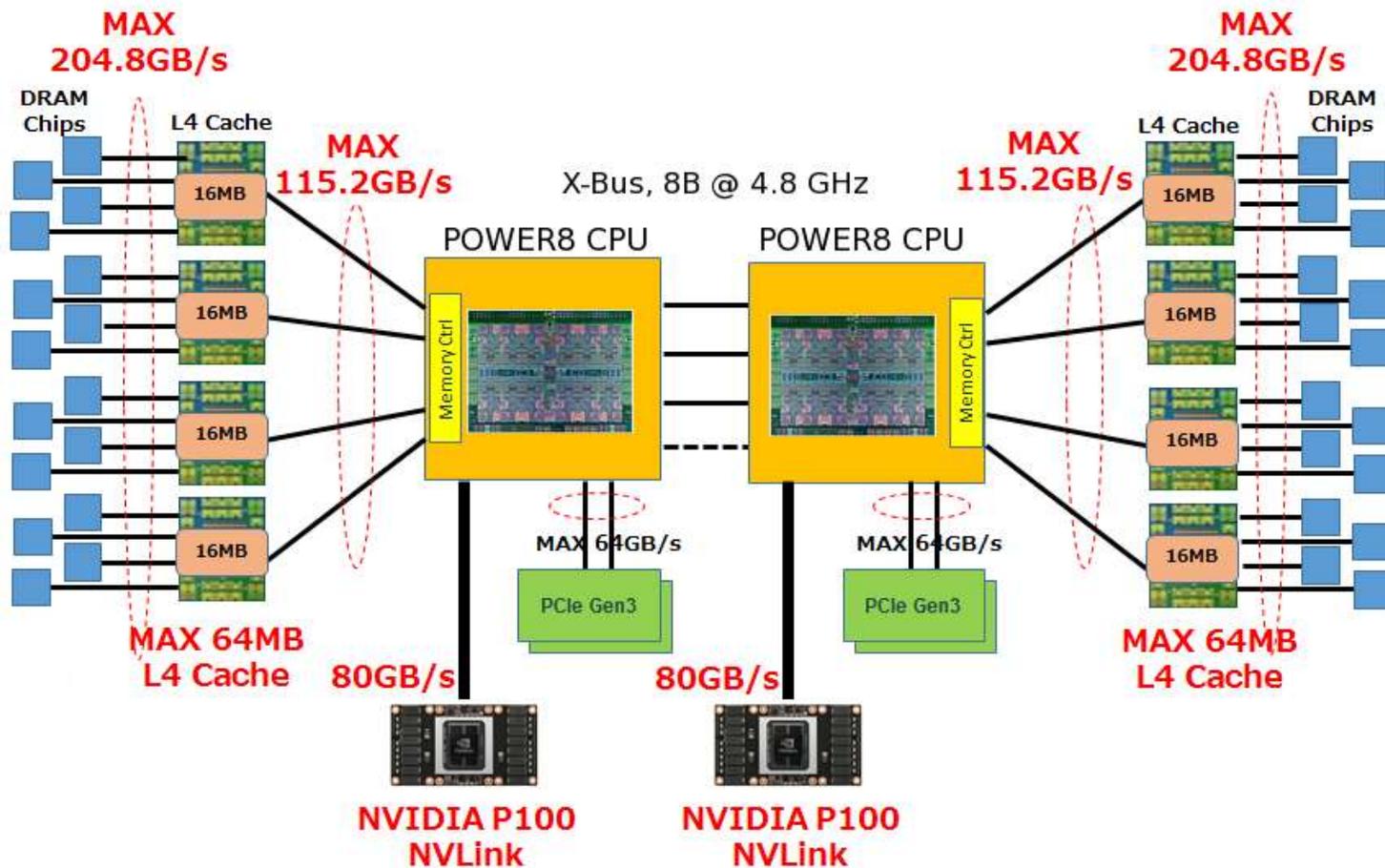
10X
Perf Over Titan

Ускорение
науки



5-10X
Application Perf Over Titan

СКЦ ДВО РАН



70% ВЫЧИСЛИТЕЛЬНОЙ НАГРУЗКИ СУПЕРКОМПЬЮТЕРОВ ПОДДЕРЖИВАЕТ GPU

GROMACS

ANSYS Fluent

Gaussian

VASP

NAMD

Simula Abaqus

WRF

OpenFOAM

ANSYS

LS-DYNA

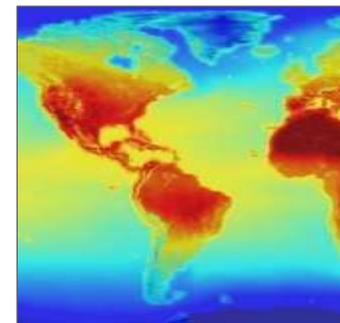
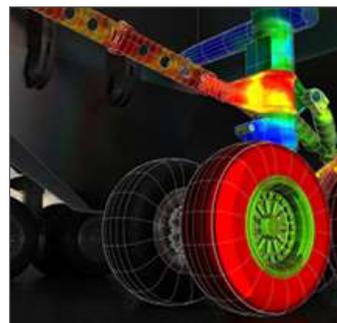
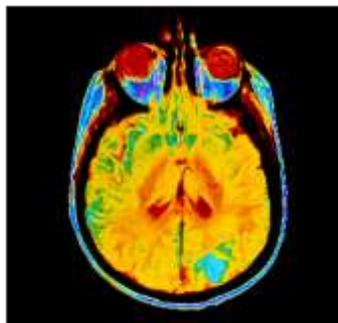
NCBI-BLAST

LAMMPS

AMBER

Quantum Espresso

GAMESS



Топ 15 HPC приложений

500+ приложений с ускорением на GPU

GPU В ОСНОВЕ ЭКОСИСТЕМЫ DEEP LEARNING

NVIDIA SDK ускоряет все основные DL-фреймворки

COMPUTER VISION

OBJECT DETECTION



IMAGE CLASSIFICATION



SPEECH & AUDIO

VOICE RECOGNITION

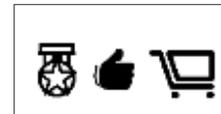


LANGUAGE TRANSLATION



NATURAL LANGUAGE PROCESSING

RECOMMENDATION ENGINES



SENTIMENT ANALYSIS



DEEP LEARNING FRAMEWORKS

Caffe



Chainer



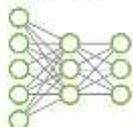
mxnet

PYTORCH theano

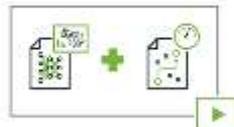


NVIDIA DEEP LEARNING SDK

cuDNN



TensorRT



DeepStream SDK



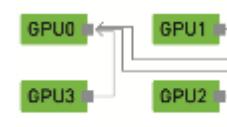
cuBLAS



cuSPARSE



NCCL



NVIDIA TESLA V100

Гигантский прорыв для искусственного интеллекта и HPC

7.5 FP64 TFLOPS | 15 FP32 TFLOPS

NEW 120 Tensor TFLOPS

16GB HBM2 @ 900 GB/s

300 GB/s NVLink



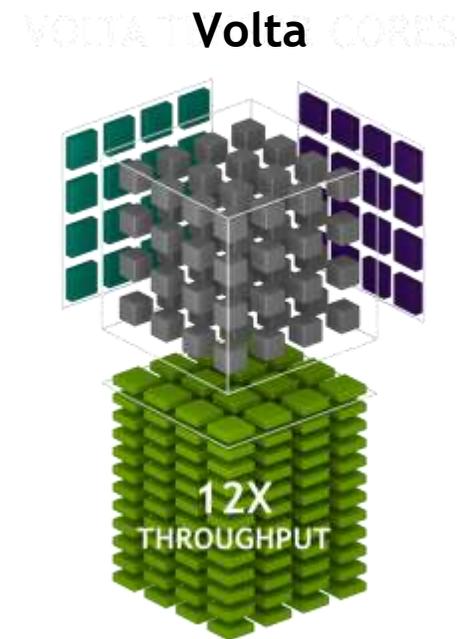
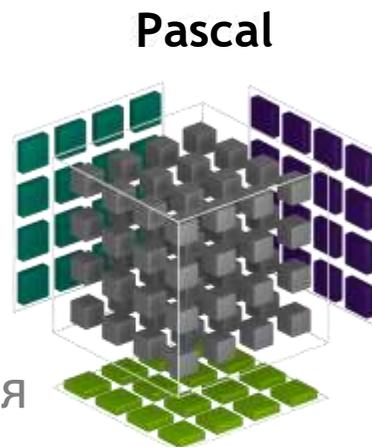
NVIDIA TESLA V100

Новые Tensor Core ядра для ИИ

Новые инструкции и форматы данных CUDA TensorOp для работы с матрицами 4x4

$$D[\text{FP32}] = A[\text{FP16}] * B[\text{FP16}] + C[\text{FP32}]$$

Оптимизированы для обучения и применения глубоких нейросетей



■ Activation Inputs ■ Weights Inputs ■ Output Results

NVIDIA DGX-1 НА БАЗЕ TESLA V100

Мощнейший инструмент для исследования в области ИИ

960 Tensor TFLOPS | 8x Tesla V100 | NVLink Hybrid Cube

Заменяет 400 традиционных серверов на задачах ИИ



NVIDIA DGX-1 НА БАЗЕ TESLA V100

Мощнейший инструмент для исследования в области ИИ

960 Tensor TFLOPS | 8x Tesla V100 | NVLink Hybrid Cube

Заменяет 400 традиционных серверов на задачах ИИ

60 ТФлопс FP64 - достаточно для Top50

3U в стойке, 3.2 кВт

Интерконнект 4 IB EDR + Dual 10 GbE



NVIDIA DGX-1 НА БАЗЕ TESLA V100

Мощнейший инструмент для исследования в области ИИ

960 Tensor TFLOPS | 8x Tesla V100 | NVLink Hybrid Cube

Заменяет 400 традиционных серверов на задачах ИИ

60 ТФлопс FP64

3U в стойке, 3.2 КВт

Интерконнект 4 IB EDR + Dual 10 GbE

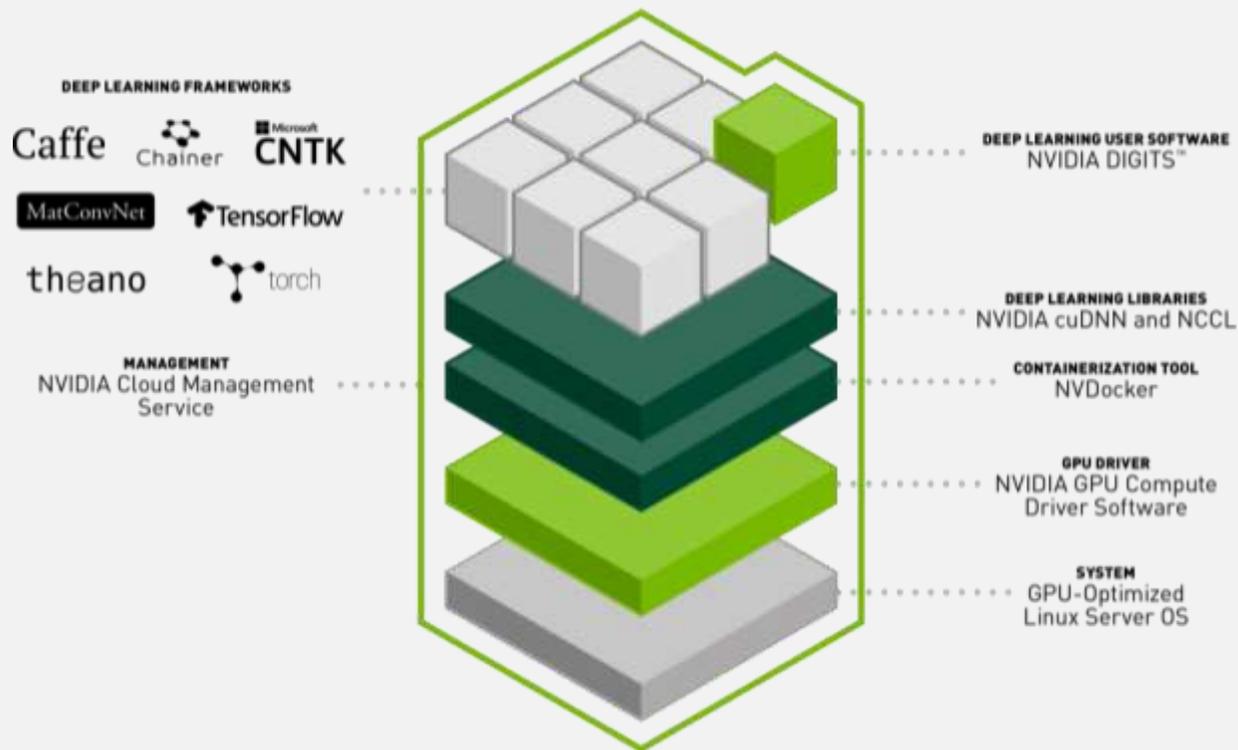
Срок поставки: 1 месяц

Пуск в эксплуатацию: 1 день



ПРОГРАММНЫЙ СТЕК DGX

Полностью интегрированная платформа для глубокого обучения



Plug-and-play - производительность здесь и сейчас

Поддержка всех фреймворков глубокого обучения

Оптимизация всех компонентов решения для максимальной производительности

Постоянные обновления из облака

Поддержка нескольких, одновременно работающих фреймворков, благодаря контейнеризации

Командная работа

Простота эксплуатации



ПОЛЬЗОВАТЕЛИ NVIDIA DGX-1



DLI TEACHING KIT

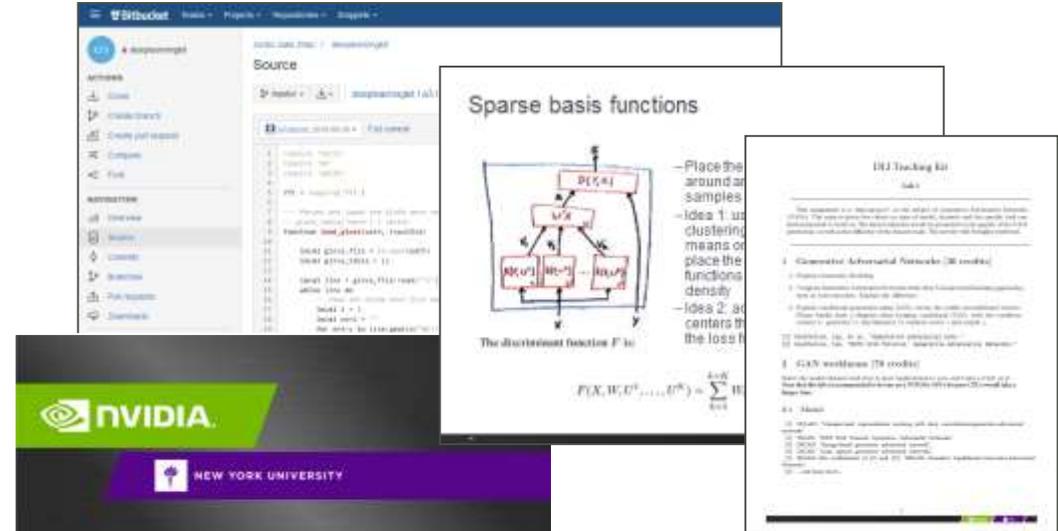


DEEP
LEARNING
INSTITUTE

Свободный доступ для преподавателей!

- Основан на курсе Deep Learning by Yann LeCun's at NYU
- Полный набор учебных материалов
 - Лекции и презентации
 - Лабораторные и примеры с решениями
 - Тесты и экзаменационные вопросы
- GPU ресурсы
 - Доступ к облачным ресурсам NVIDIA DLI Qwiklab для студентов
 - До \$125 Amazon Web Services (AWS) GPU instances для студентов и \$200 для преподавателя в семестр

<https://developer.nvidia.com/teaching-kits>



Курсы читаются в:



Лабораторные работы



DEEP
LEARNING
INSTITUTE

Более 20 прикладных лабораторных работ по deep learning

Некоторые темы:

- . Object detection
- . Image classification
- . Image creation using Generative Adversarial Networks
- . Image segmentation
- . Signal processing
- . Medical image analysis
- . Medical image segmentation
- . Deep learning for genomics using DragoNN with Keras and Theano
- . Trading strategy for finance using LSTMs

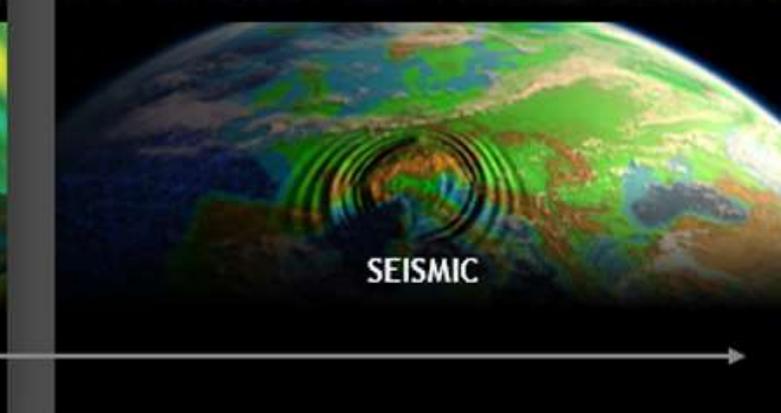
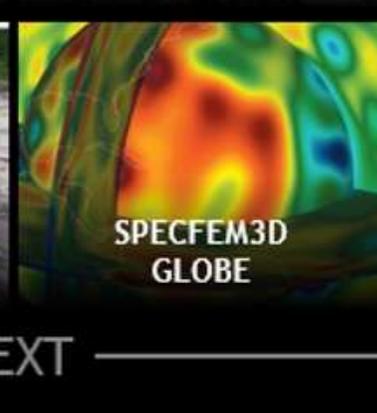
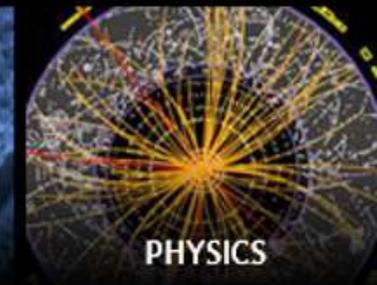
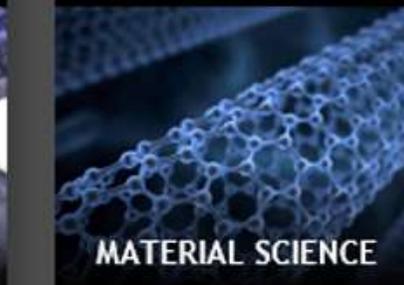
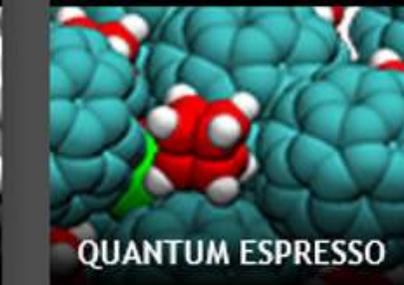
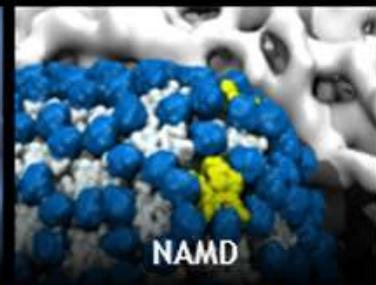
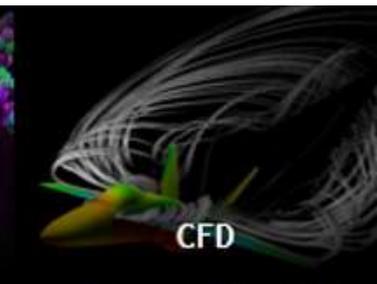
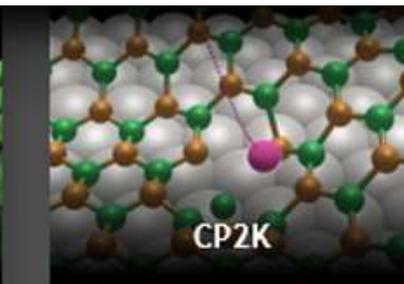
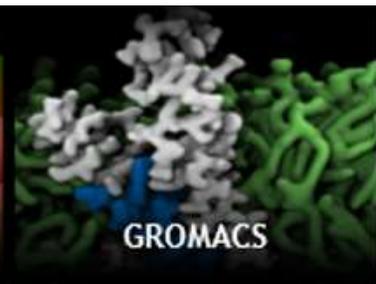
NVIDIA GPU CLOUD ДЛЯ HPC

ОБЛАЧНЫЙ РЕЕСТР КОНТЕЙНЕРОВ С HPC ПРИЛОЖЕНИЯМИ

- Готовые собранные контейнеры NVDocke
- Оптимизировано для GPU-ускоренных систем
- Постоянные обновления
- Доступны свободно
- Регистрация на nvidia.com/gpu-cloud



НРС ПРИЛОЖЕНИЯ В NVIDIA GPU CLOUD



NEXT



NVIDIA GPU CLOUD ДЛЯ ВИЗУАЛИЗАЦИИ

УНИФИЦИРОВАННАЯ ВИЗУАЛИЗАЦИЯ БОЛЬШИХ ОБЪЕМОВ ДАННЫХ

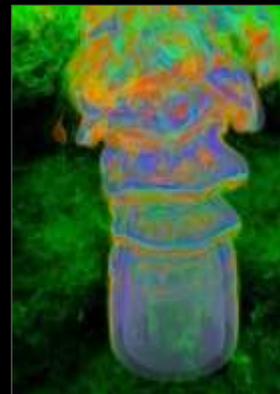
Объемная визуализация

Физически корректный рендеринг

Качественные изображения

Интеграция с ParaView

Ранний доступ на nvidia.com/gpu-cloud



ParaView with
NVIDIA IndeX

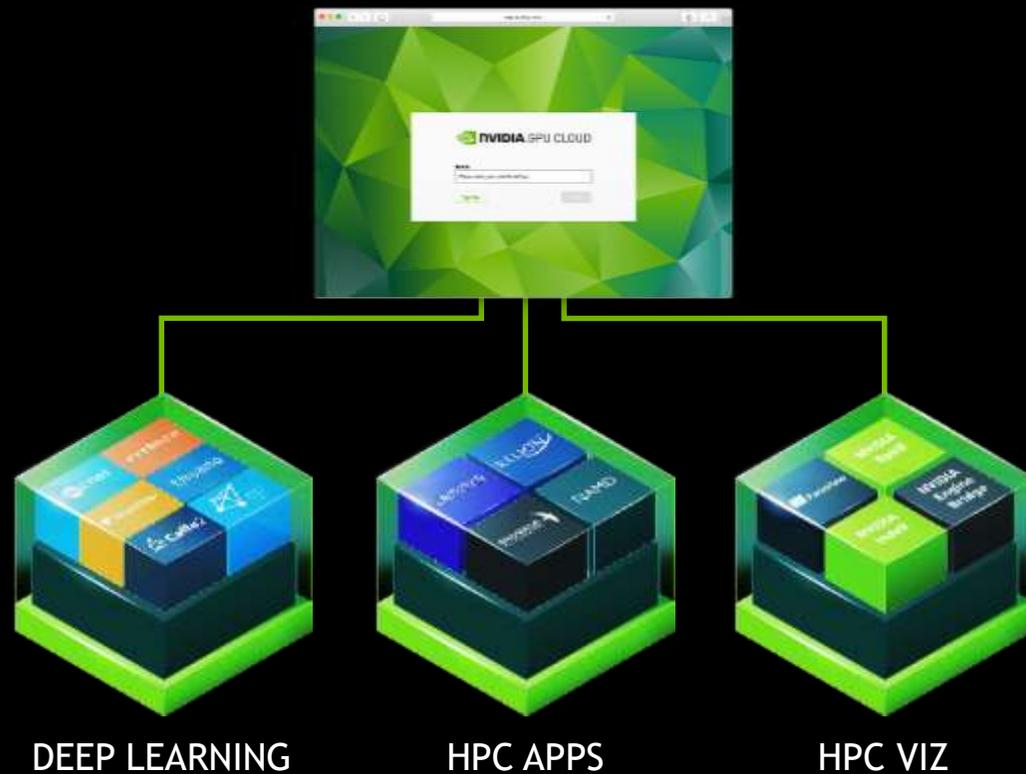


ParaView with
NVIDIA OptiX



ParaView with
NVIDIA Holodeck

NVIDIA GPU CLOUD УПРОЩАЕТ РАБОТУ С HPC И DL



Сегодня каждый вычислительный центр должен иметь инструменты для ИИ:

1. систему DGX-1 или процессоры Tesla
2. курсы по глубоким нейросетям.

Антон Джораев, adzhoraev@nvidia.com

