

Introduction to HPC

CSCI-4850/5850 High-Performance Computing

Summer 2021 at GIST

Tae-Hyuk (Ted) Ahn

Department of Computer Science
Program of Bioinformatics and Computational Biology
Saint Louis University



**SAINT LOUIS
UNIVERSITY™**

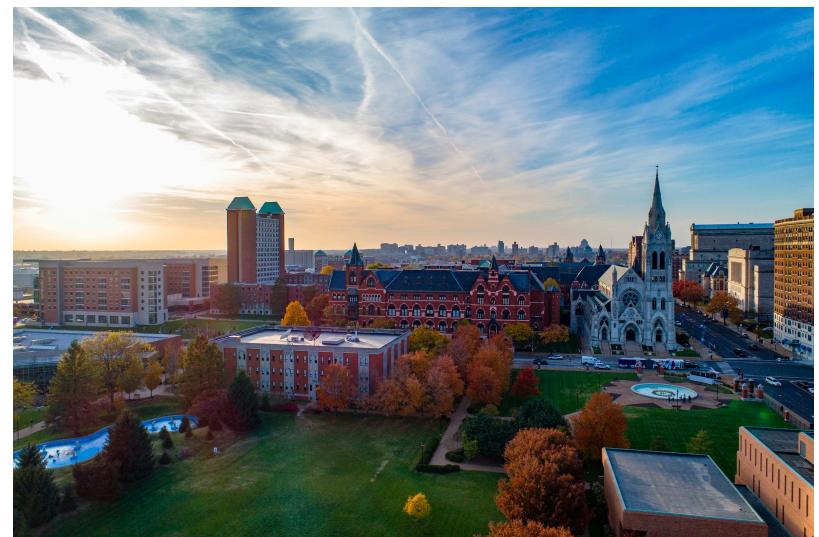
— EST. 1818 —

Welcome

- Welcome to **High-Performance Computing**

- Let me introduce!

- Graduated EE at 2000
- Worked at Samsung SDS from 2000 – 2004
- Worked as a Math teacher before going abroad
- July 2005 ~ staying at US
 - ECE, Northwestern University (Chicago, IL) 2005 – 2007
 - CS, Virginia Tech (Blacksburg, VA) 2007 – 2012
 - Modeling and Simulation Cell Cycle Model (in parallel)
 - RNA-Seq and some bioinformatics projects
 - Sandia National Lab HPC Intern (Livermore, CA) 2010
 - Pfizer Inc. HPC Intern (Groton, CT) 2011
 - Oak Ridge National Lab CS & Bioinformatics Group (2012 – 2015)
 - Saint Louis University (2015 – present)



Welcome

- Tell me about you for me to memorize your name/face quickly!
 - What is your name, major, and year?
 - Do you have something that I would like to remember you easily?

Survey (to understand your background)

- Go to Week 1 and complete the “Assignment 1” survey in the class time (15 mins)
- Let’s check the statistics together!

Let us go over the syllabus

- The syllabus could be updated again, so please check the syllabus often.

Topical Outline

Main topics include basic HPC hardware architectures, parallel programming languages, new trends of HPC, and more. The contents of the course provide a balance of theoretical and practical aspects in parallel computing. A student of this course is expected to develop the right skills to design parallel applications and to effectively use modern HPC platforms.

- Introduction to HPC and parallel programming, Serial code optimization
- Shared-memory parallel programming with OpenMP
- Distributed-memory parallel programming with MPI
- GPU programming with CUDA
- Cloud Computing and Apache Spark

Student Learning Outcomes

At the completion of this course, students will be able to:

- Understand fundamental concepts and new trends of high-performance computing
- Optimize software to take advantage of processor feature
- Design, implement, and analyze programs via OpenMP for shared-memory system
- Design, implement, and analyze programs via MPI for distributed-memory system
- Understand the concept and implement simple GPU programs with CUDA
- Learn cutting-edge big-data cloud-computing framework using Apache Spark
- Learn ability to decide optimization or parallelize of existing applications

Course Textbook and Resources

- Textbook: No textbook is required for this course. Assigned materials will be posted on class website. Optional reference books are as below:
 - An Introduction to Parallel Programming by Peter Pacheco (2011, ISBN:978-0123742605).
 - Introduction to High Performance Computing for Scientists and Engineers by Georg Hager and Gerhard Wellein (2010, ISBN-13: 978-1439811924)
 - Introduction to Parallel Computing by Ananth Grama, George Karypis, Vipin Kumar, and Anshul Gupta (2003, ISBN:978-0201648652).
 - Principles of Parallel Programming by Calvin Lin, and Larry Snyder (2008, ISBN:978-0321487902).
 - Professional CUDA C Programming by John Cheng, Max Grossman, Ty McKercher (ISBN: 978-1-118-73932-7)
 - Spark in Action, Second Edition: Covers Apache Spark 3 with Examples in Java, Python, and Scala by Jean-Georges Perrin (2020, ISBN-13: 978-1617295522)
 - Learning Spark: Lightning-Fast Data Analytics by Jules S. Damji (2020, ISBN-13: 978-1492050049)

Computing Resource, Terminal, Linux, and Shell

- After I hear your background by survey, I will decide what resources will be provided.

- I strongly recommend you to install Jupyter Lab in your laptop or your desktop.
- <https://jupyter.org/>
- If you cannot get started, watch the YouTube video “Jupyter Notebook Tutorial: Introduction, Setup, and Walkthrough” (<https://www.youtube.com/watch?v=HW29067qVWk>)

1. What is the Jupyter Notebook?

In this page briefly introduce the main components of the *Jupyter Notebook* environment. For a more complete overview see [References](#).

Contents

- [What is the Jupyter Notebook?](#)
 - [Notebook document](#)
 - [Jupyter Notebook App](#)
 - [kernel](#)
 - [Notebook Dashboard](#)
 - [References](#)

1.1. Notebook document

Notebook documents (or “notebooks”, all lower case) are documents produced by the [Jupyter Notebook App](#), which contain both computer code (e.g. python) and rich text elements (paragraph, equations, figures, links, etc...). Notebook documents are both human-readable documents containing the analysis description and the results (figures, tables, etc..) as well as executable documents which can be run to perform data analysis.

References: Notebook documents [in the project homepage](#) and [in the official docs](#).

1.2. Jupyter Notebook App

The *Jupyter Notebook App* is a server-client application that allows editing and running [notebook documents](#) via a web browser. The *Jupyter Notebook App* can be executed on a local desktop requiring no internet access (as described in this document) or can be installed on a remote server and accessed through the internet.

Jupyter Lab

← → ⌂ ⌄ 🔒 jupyterlab.readthedocs.io/en/stable/

 JupyterLab
stable

Search docs

GETTING STARTED

- Overview
- Installation
- Starting JupyterLab
- Reporting an issue
- Frequently Asked Questions (FAQ)
- JupyterLab Changelog

USER GUIDE

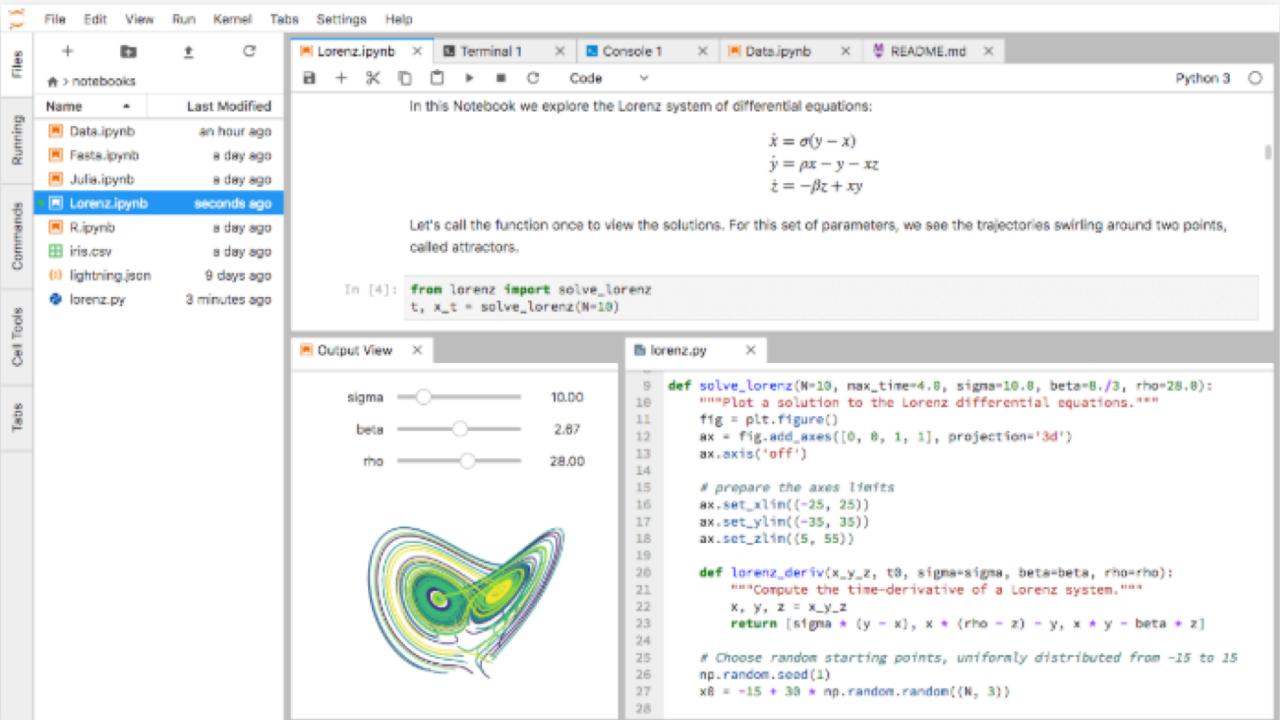
- The JupyterLab Interface
- JupyterLab URLs
- Working with Files
- Text Editor
- Notebooks
- Code Consoles
- Terminals
- Managing Kernels and Terminals
- Command Palette

Docs » JupyterLab Documentation

 Jupyter |  Edit on GitHub

JupyterLab Documentation

JupyterLab is the next-generation web-based user interface for Project Jupyter. [Try it on Binder](#).
JupyterLab follows the [Jupyter Community Guides](#).



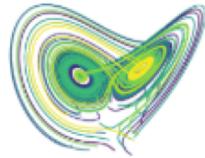
In this Notebook we explore the Lorenz system of differential equations:

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= -\beta z + xy\end{aligned}$$

Let's call the function once to view the solutions. For this set of parameters, we see the trajectories swirling around two points, called attractors.

```
In [4]: from lorenz import solve_lorenz
t, x_t = solve_lorenz(N=10)
```

sigma: 10.00
beta: 2.87
rho: 28.00



```
def solve_lorenz(N=10, max_time=4.0, sigma=10.0, beta=8./3, rho=28.0):
    """Plot a solution to the Lorenz differential equations."""
    fig = plt.figure()
    ax = fig.add_axes([0, 0, 1, 1], projection='3d')
    ax.axis('off')

    # prepare the axes limits
    ax.set_xlim((-25, 25))
    ax.set_ylim((-35, 35))
    ax.set_zlim((5, 55))

    def lorenz_deriv(x_y_z, t0, sigma=sigma, beta=beta, rho=rho):
        """Compute the time-derivative of a Lorenz system."""
        x, y, z = x_y_z
        return [sigma * (y - x), x * (rho - z) - y, x * y - beta * z]

    # Choose random starting points, uniformly distributed from -15 to 15
    np.random.seed(1)
    x0 = -15 + 30 * np.random(N, 3)
```

Tentative Class Environments

- Terminal, Linux, and Shell Programming
- Mainly C++, Some Python

Grading

- **Homework and Unit Exam Assignments (70%)**
 - There will be written and programming assignments in a variety of format assignments during the course. Maximum points of the assignments could be different. Check each assignment submission due carefully and submit it on time.
- **Final Report (20%)**
 - You will be asked to write and submit a final report of 2~3 pages.
 - If you have been working on a specific research project,
 - Provide research background.
 - Survey how HPC is used in your research area.
 - Propose your approach and solution to accelerate your research.
 - If you do not have a research topic yet,
 - Search your interesting topic (e.g., Covid19 + HPC) literatures and review those papers intensively.
 - You can review one paper or multiple papers on the topic.
 - You also have to provide your thoughts and critiques.
- **Attendance and Activities (10%)**
 - If you cannot join synchronous lectures, you should contact me in advance to provide a reason for the absence. If not, -0.5 point for each absence.

Welcome to HPC course!

- Welcome again!!!
- I will try to provide lots of materials/opportunities to you!!!

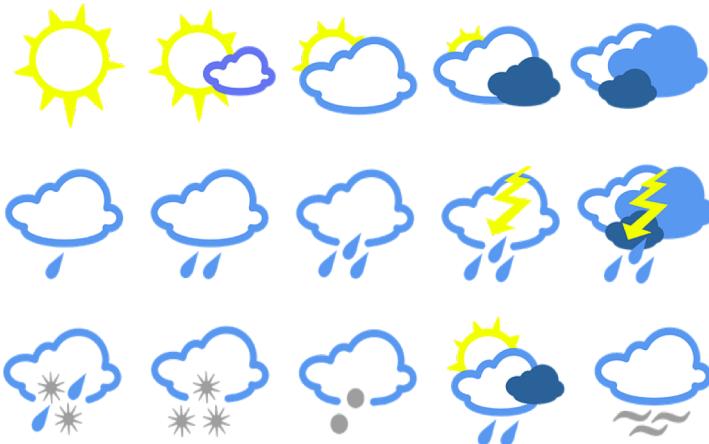
What is HPC?

- **High-performance computing (HPC)** is the use of parallel processing for running advanced application programs efficiently, reliably and quickly.
- **High performance computing (HPC)** is a class of applications and workloads that perform computationally intensive operations across multiple resources. – Intel
- “**High-Performance Computing** most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business” – <https://www.usgs.gov/>
- “**High-Performance Computing (HPC)**” is the application of "supercomputers" to computational problems that are either too large for standard computers or would take too long. A desktop computer generally has a single processing chip, commonly called a CPU. A HPC system, on the other hand, is essentially a network of nodes, each of which contains one or more processing chips, as well as its own memory. - <https://www.nics.tennessee.edu/>

What is HPC?

- Definition depends on individual person
 - HPC is when I care how fast I get an answer
- Thus HPC can happen on:
 - A workstation, desktop, laptop, smartphone!
 - A supercomputer
 - A Linux/MacOS/Windows/... cluster
 - A grid or a cloud
 - Cyberinfrastructure = any combination of the above
- HPC also means High-**Productivity** Computing

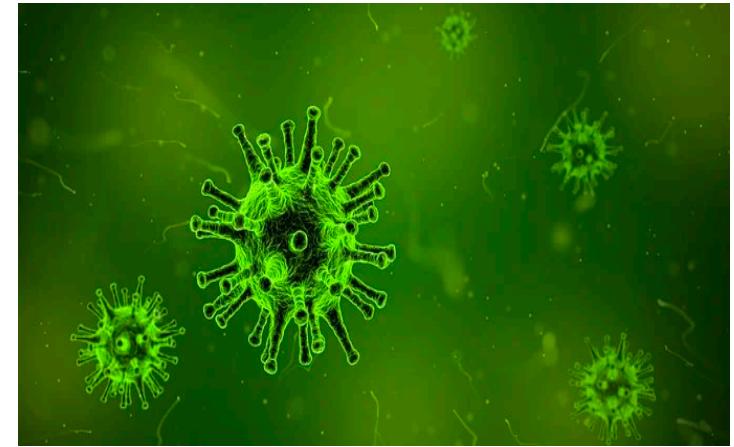
Welcome to HPC World!



Climate



Finance



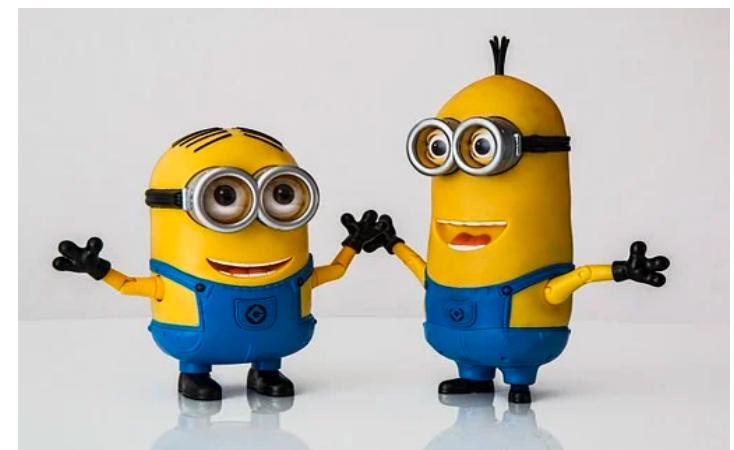
Medical



Bioinformatics



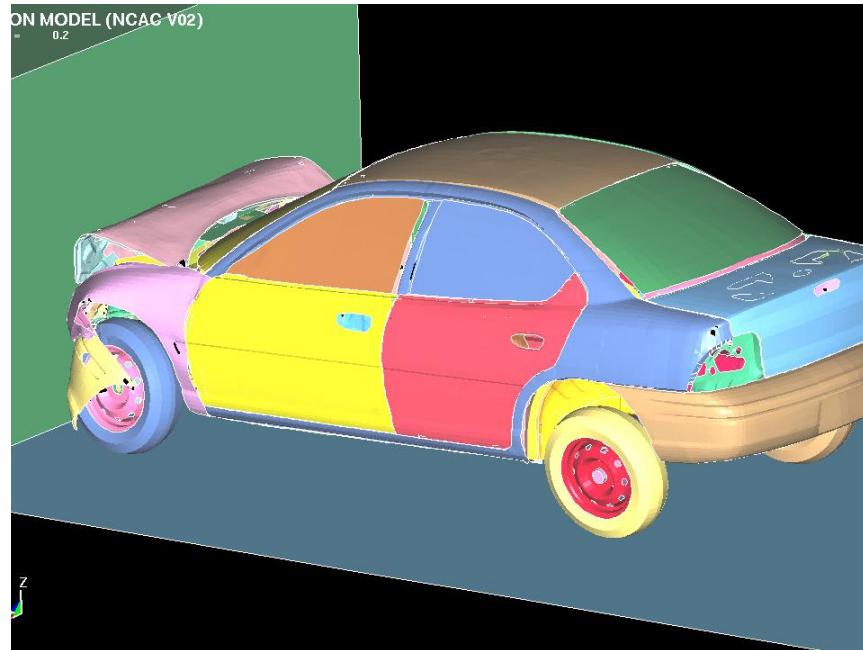
Physics



Entertainment

Why do we need HPC?

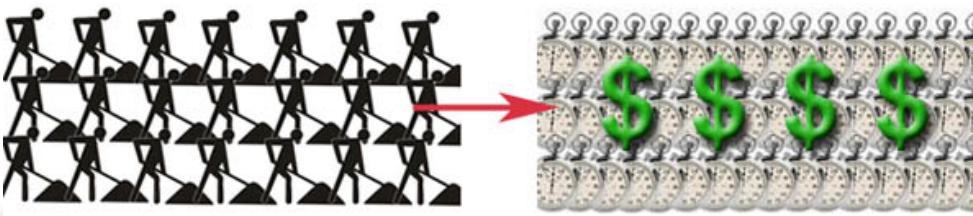
Save Time and Money



\$500,000

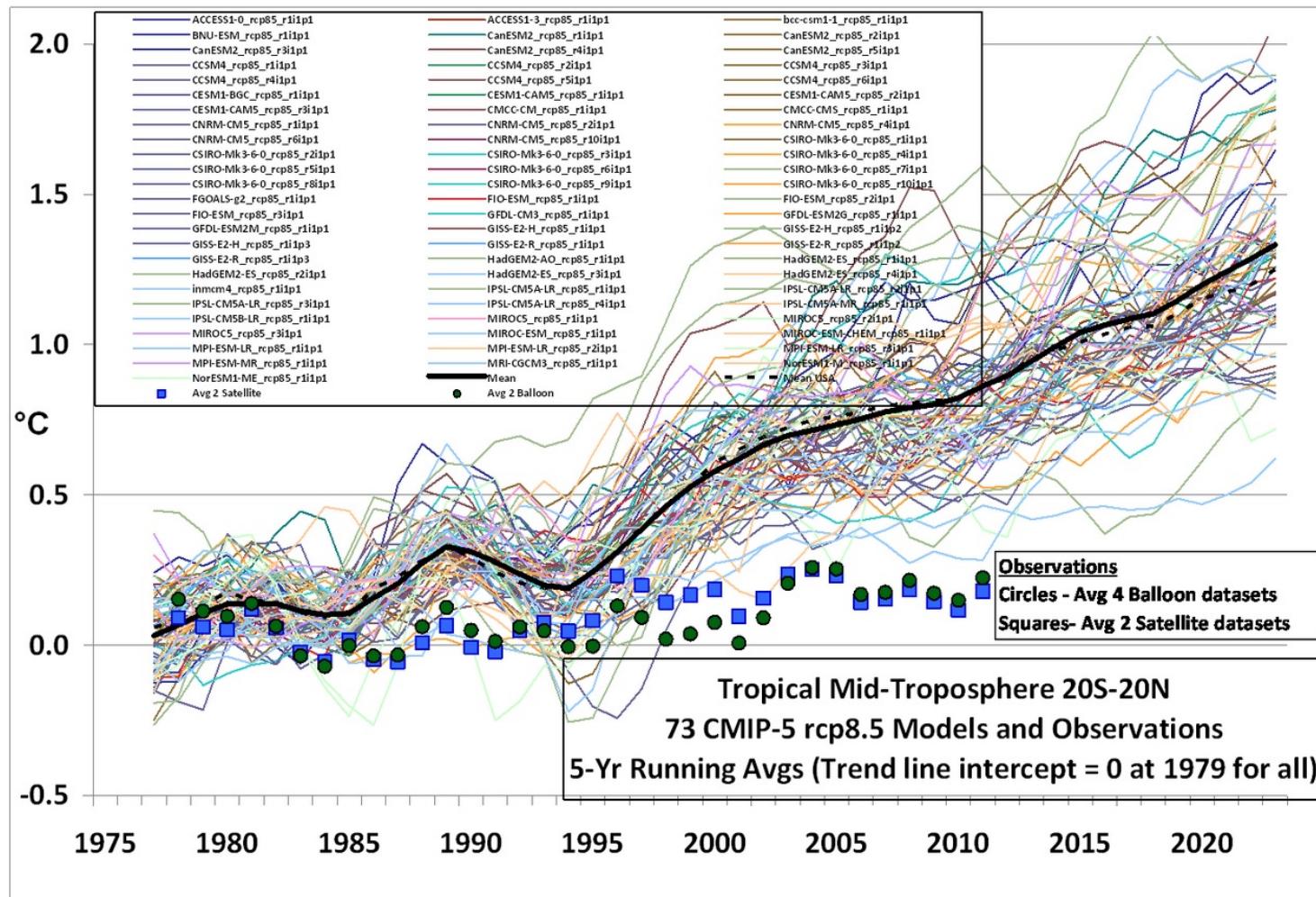
VS

\$10



Why do we need HPC?

Human cannot calculate, but computer can!



Grand Challenge Problems

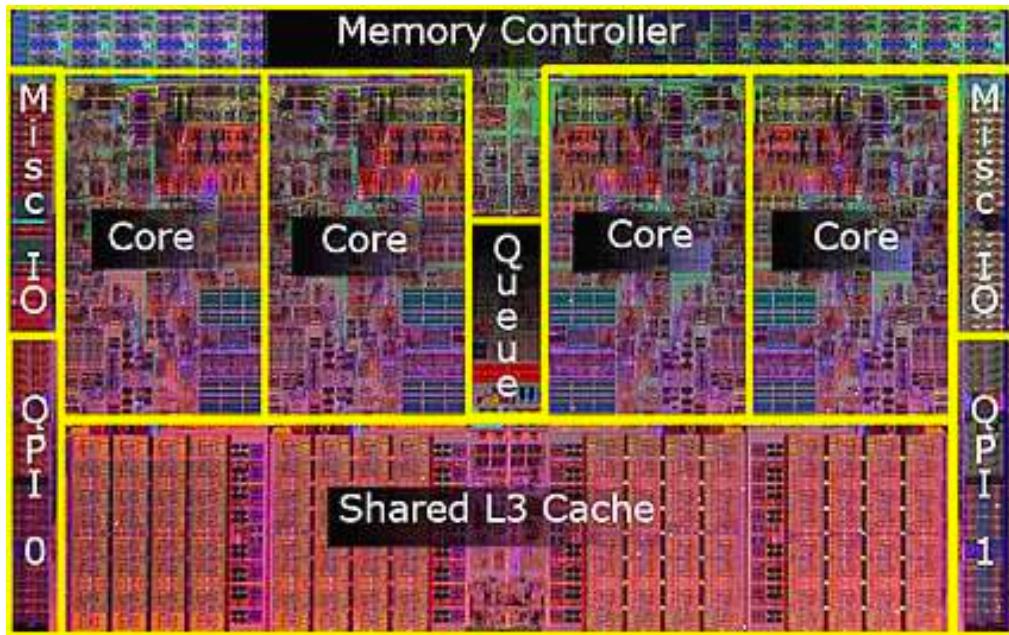
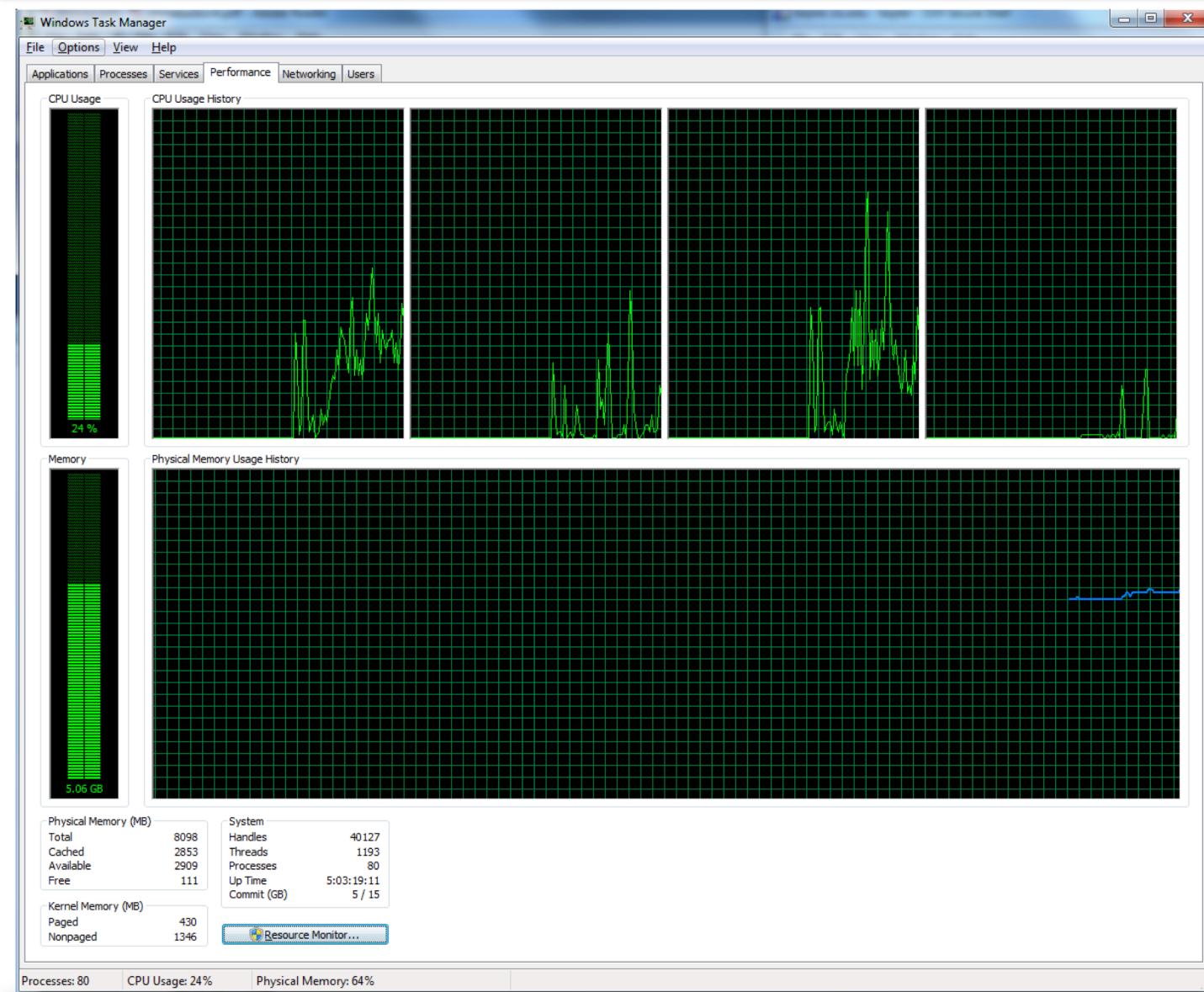
Big problems

- A “**Grand Challenge**” problem is a problem that cannot be solved in a reasonable amount of time with today’s computers
- Examples of Grand Challenge problems:
 - Human health and disease
 - Applied Fluid Dynamics
 - Meso- to Macro-Scale Environmental Modeling
 - Ecosystem Simulations
 - Biomedical Imaging and Biomechanics
 - Molecular Design and Process Optimization
 - Fundamental Computational Sciences
 - Nuclear power and weapons simulations

Why would HPC matter to you?

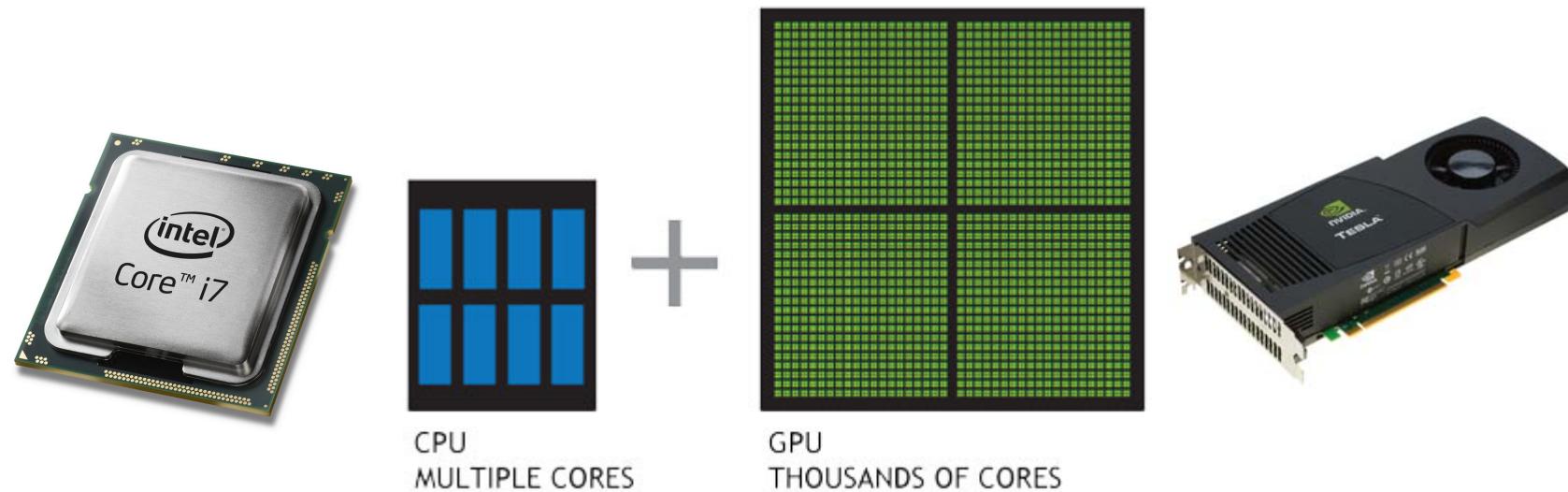
- Scientific computing is becoming more important in many research disciplines
- Problems become more complex, need teams of researchers with diverse expertise
- Scientific (HPC) application development often limited by lack of training
- More knowledge about HPC leads to more effective use of HPC resources and better interactions with (computational) colleagues

Your Laptop is HPC?



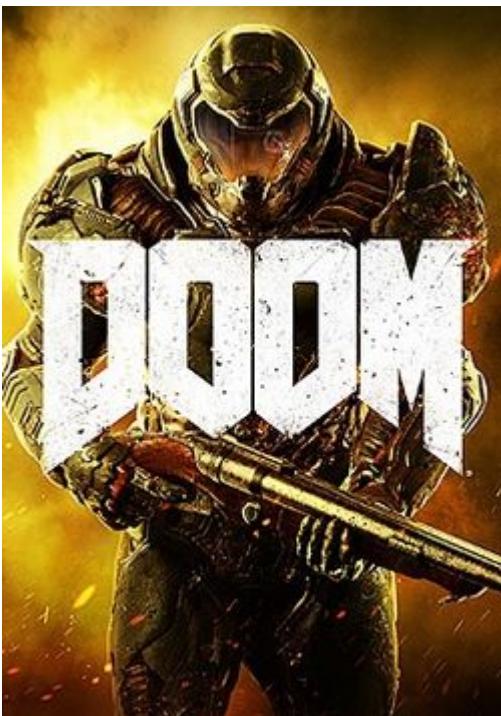
New Trend of Desktops/Laptops

- Two major trends:
 - Through the application of multicore computing. Processor architects focus on throughput, not clock speed, to improve performance.
 - Access to widely available **graphic processing units (GPUs)** for general processing.



GPU

- GPU: Graphics Processing Unit
 - Designed for real-time graphics
 - Present in almost every PC
 - Increasing realism and complexity



NVIDIA GPU

The Best Budget Graphics Cards in 2020 (by my quick and inaccurate search, I am not playing game anymore)

SHOP ALL GEFORCE RTX

RTX 30 SERIES RTX 20 SERIES



GEFORCE RTX 3090

GEFORCE RTX 3080 Ti

GEFORCE RTX 3080

GEFORCE RTX 3070 Ti

[SEE ALL BUYING OPTIONS](#)

[SEE ALL BUYING OPTIONS](#)

[SEE ALL BUYING OPTIONS](#)

[SEE ALL BUYING OPTIONS](#)

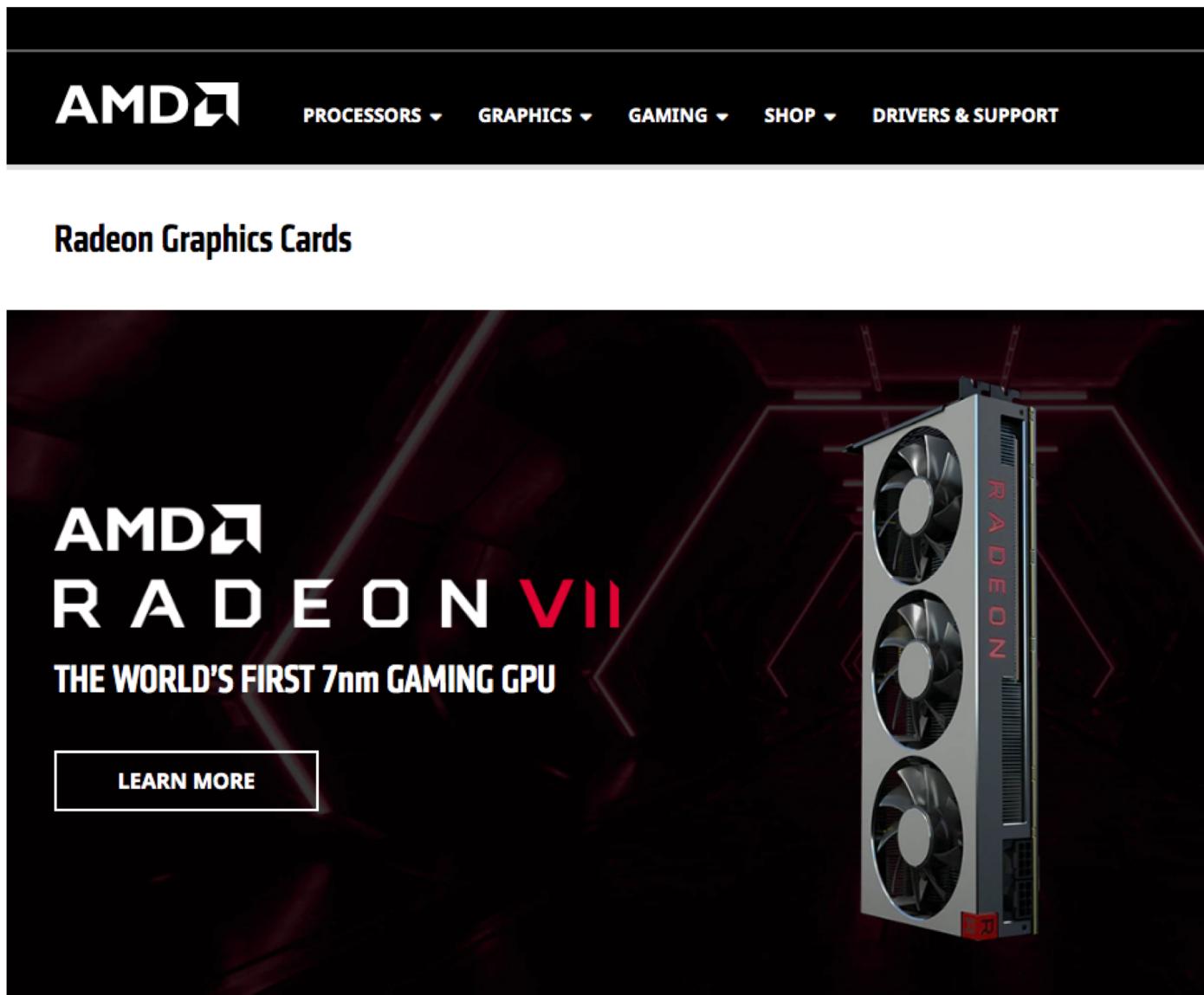
[LEARN MORE >](#)

[LEARN MORE >](#)

[LEARN MORE >](#)

[LEARN MORE >](#)

AMD Radeon



The image shows a screenshot of the AMD Radeon website. At the top, there is a black navigation bar with the AMD logo and links for PROCESSORS, GRAPHICS, GAMING, SHOP, and DRIVERS & SUPPORT. Below the navigation bar, the text "Radeon Graphics Cards" is displayed. On the left side of the main content area, the text "AMD RADEON VII" is prominently shown, followed by "THE WORLD'S FIRST 7nm GAMING GPU". A "LEARN MORE" button is located at the bottom left. On the right side, there is a large image of the AMD Radeon VII graphics card, which has three large fans and a red "RADEON" logo on its backplate.

- My MacBook Pro has Radeon Pro 560



The image shows a screenshot of the macOS High Sierra System Report interface. At the top, it says "macOS High Sierra Version 10.13.6". Below that, it lists system specifications: "MacBook Pro (15-inch, 2017)", "Processor 2.9 GHz Intel Core i7", "Memory 16 GB 2133 MHz LPDDR3", "Graphics Radeon Pro 560 4096 MB Intel HD Graphics 630 1536 MB", and "Serial Number C02WH0NFHTDF". At the bottom, there are two buttons: "System Report..." and "Software Update...".

GPU compute built for deep learning

GPU cloud, workstations, servers, and laptops built for deep learning. Speed up PyTorch, TensorFlow, Keras, and save up to 90%.

Talk to an engineer



Deep Learning Laptop



GPU Workstation



GPU Cloud



GPU Cluster



Quadro GPU Server



Tesla GPU Server

New! Quadro RTX A6000, RTX 3090, RTX 3080, RTX 3070 pre-order >

Deep learning workstation with up to 4 GPUs

NVIDIA RTX 3090, RTX 3080, RTX 3070, Titan RTX, Quadro RTX 5000, RTX 6000, RTX 8000, and RTX A6000 options. Pre-installed with Ubuntu, TensorFlow, PyTorch, Keras, CUDA, and cuDNN.

Customize now

Technical specifications

GPU	Up to 4x NVIDIA GPUs Choose from RTX 30 series (Ampere), Titan RTX, Quadro RTX 8000, and RTX 6000
Processor	AMD Ryzen or Intel Core i9 Configurable up to 64 cores, 128 threads, and 256 MB cache
Memory	Up to 256 GB Fits up to eight 32 GB DIMMs at 3200 MHz



From \$5,486.00

Free 30-day returns

Processor

AMD Threadripper 3960X

24 cores, 3.80 GHz, 128 MB cache

\$5,486.00

 Apply for a discount

GPUs

2x RTX 3070 (pre-order)

Memory

64 GB

Operating System Drive

1 TB SSD (NVMe)

- \$336

2 TB SSD (NVMe)



Workstation, Cluster, Supercomputer?

- Most **workstation** computers today are parallel workstations
 - multi-core processors
- Running Linux OS (or MacOS X) allows programming like traditional Unix workstation
- Software to use multi-cores/threads
 - Pthreads, OpenMP, Multiprocessing
- All processors have access to all memory
 - Uniform memory access (UMA):
 - 1 memory pool for all, same speed for all
 - Non-uniform memory access (NUMA):
 - multiple pools, speed depends on “distance”

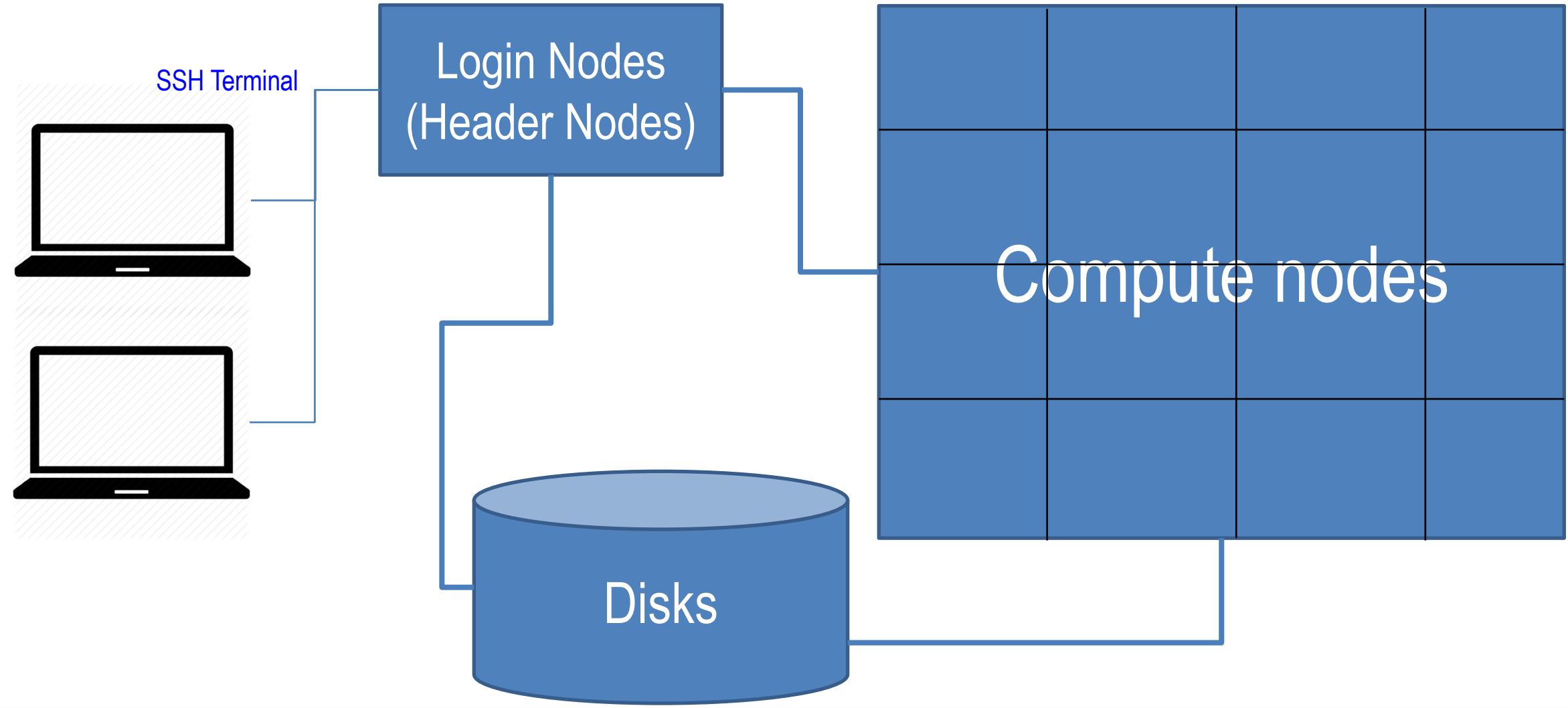


Workstation, **Cluster**, Supercomputer?

- A **cluster** needs:
 - Several computers, **nodes**, often in special cases for easy mounting in a rack
 - One or more networks (interconnects) to hook the nodes together
 - Software that allows the nodes to communicate with each other (e.g. MPI)
 - Software that reserves resources to individual users
- A cluster is: all of those components working together to form one big computer



Typical HPC (Cluster/Server) System Layout



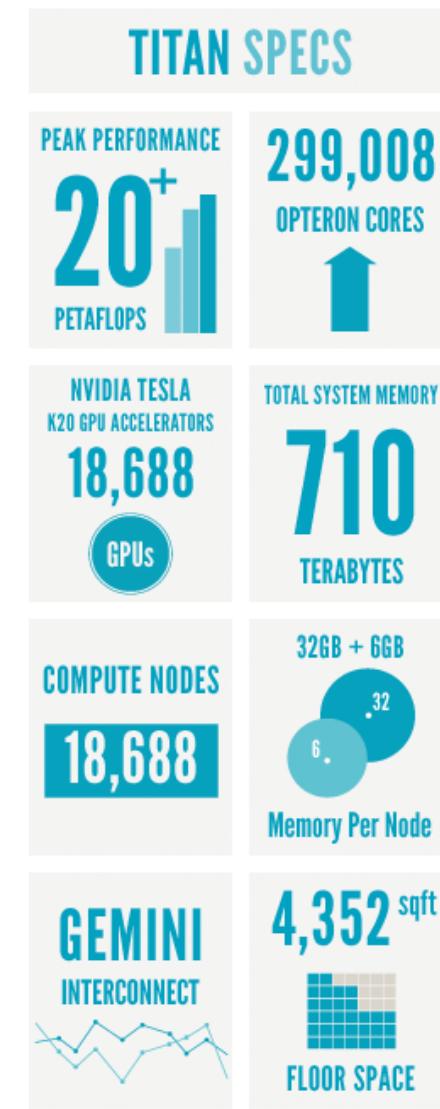
Workstation, Cluster, Supercomputer?

- A **supercomputer** is a computer with a high-level computational capacity compared to a general-purpose computer.
 - The most visible manifestation of HPC
 - Programs run on the fastest and largest computers in the world (Top500 List, <https://www.top500.org/>)
 - Performance of a supercomputer is measured in tera scale floating-point operations per second (Tflops)



Rank	System	Cores	Rmax [TFlop/s]	Rpeak [TFlop/s]	Power [kW]
1	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
3	Sierra - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
4	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371

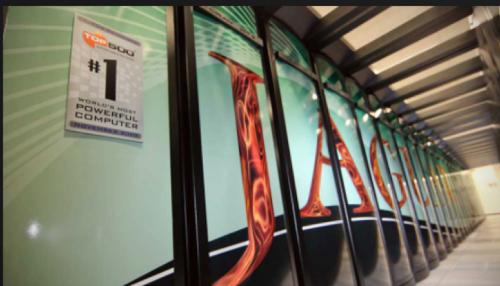
³⁴Oak Ridge National Lab Supercomputing Center (2012-2015)



Oak Ridge National Lab Supercomputing Center

WORLD LEADING SYSTEMS

Oak Ridge National Laboratory has decades of experience in delivering, operating, and conducting research on world-leading supercomputers. Since 2005, Oak Ridge National Laboratory has deployed Jaguar, Titan, and Summit, each the world's fastest computer in its time. Frontier will leverage ORNL's extensive experience and expertise in GPU-accelerated computing to become the US Department of Energy's next record-breaking supercomputer when it debuts in 2021.



JAGUAR

No. 1 in 2009, 2010



TITAN

No. 1 in 2012



SUMMIT

No. 1 in 2018



FRONTIER

Coming in 2021

Cloud Computing

- "Cloud Computing", by definition, refers to the on-demand delivery of IT resources and applications via the Internet with pay-as-you-go pricing.
- Cloud Computing provides a simple way to access servers, storage, databases and a broad set of application services over the Internet.
- Cloud Computing providers such as Amazon Web Services own and maintain the network-connected hardware required for these application services, while you provision and use what you need via a web application.



What is Parallel Computing?

- Parallel computing is the use of multiple processing entities in combination to solve a single problem.
- It is not always obvious that a parallel algorithm has benefits, unless we want to do things ...
 - faster: doing the same amount of work in less time
 - bigger: doing more work in the same amount of time
- Both of these reasons can be argued to produce better results, which is the only meaningful outcome of program parallelization

Parallel Processing with Big Data

- Discrepancy between the explosive growth rate in data volumes and the improvement trends in processing and memory access speeds necessitates that parallel processing be applied to the handling of extremely large data sets.



- Google AlphaGo

Configuration and strength ^[63]			
Versions	Hardware	Elo rating	Matches
AlphaGo Fan	176 GPUs, ^[52] distributed	3,144 ^[51]	5:0 against Fan Hui
AlphaGo Lee	48 TPUs, ^[52] distributed	3,739 ^[51]	4:1 against Lee Sedol
AlphaGo Master	4 TPUs, ^[52] single machine	4,858 ^[51]	60:0 against professional players; Future of Go Summit
AlphaGo Zero	4 TPUs, ^[52] single machine	5,185 ^[51]	100:0 against AlphaGo Lee 89:11 against AlphaGo Master
AlphaZero	4 TPUs, single machine	N/A	60:40 against AlphaGo Zero

The screenshot shows a web browser displaying the DeepMind blog post titled "AlphaGo Zero: Learning from scratch". The URL is https://deepmind.com/blog/alphago-zero-learning-scratch/. The page features a dark background with a faint image of a Go board. At the top, there's a navigation bar with links to Home, Research, Applied, News & Blog, About Us, and Careers. A search icon is also present. The main title "AlphaGo Zero: Learning from scratch" is centered above a detailed text section. Below the text, there's a large image of a Go board with stones and a small banner at the bottom right that says "Starting from scratch".

Artificial intelligence research has made rapid progress in a wide variety of domains from speech recognition and image classification to genomics and drug discovery. In many cases, these are specialist systems that leverage enormous amounts of human expertise and data.

However, for some problems this human knowledge may be too expensive, too unreliable or simply unavailable. As a result, a long-standing ambition of AI research is to bypass this step, creating algorithms that achieve superhuman performance in the most challenging domains with no human input. In our most recent paper, published in the journal Nature, we demonstrate a significant step towards this goal.

HPC + AI... Not matured yet...

EXCLUSIVE

STAT+

IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show

By CASEY ROSS @caseymross and IKE SWETLITZ @ikeswetlitz / JULY 25, 2018



Welcome to HPC world!

- Welcome!!!
- Lots of opportunities for you!!!