Objective:

1. Overview of the area with some *tangible basic skills*.
2. Quiz to clarify some (often misunderstood) basic material with concrete and realistic examples
3. Connect other areas to provide the *Big Picture without compromising on finer details*
4. How academic and industry look at the same problem

Scope:

1. Not entirely from Course Outline (Institute) – why?
2. Use instructor's experience in large scale computing
3. Depth and breadth are limited by course duration & Instructor's expertise
4. Exercises are based on both Windows (basic platform) and Linux (VMs)

Target group:

Students with in ± Sigma range

Grading:

- Not Relative, no bell curve
- Quiz 1: 25%; Quiz 2: 25 %; Final Exam/quiz: 50%
- Lab: 30%; Lab Final Quiz: 20%; Lab Final Exam 50%
- Lab content is included for theory quizzes and exam

Text Books:

You can use the prescribed ones; the course content is not constrained by these textbooks.

Reference Material:

Class presentation, Instructor notes and reference material suggested by the instructor

Content:

1. Pre-Requisite
   - Computing and Networking
   - "Computing is Networking" versus "Networking is Computing"!
   - Some content from your course outline!
2. Foundation Material
   - Pre-Requisites
   - What *Performance* are we talking about?
   - Client-Server Architecture with *threads*
   - Parallel Processing and Parallel Algorithms.
   - Vector Processing
3. Large Scale system
   - Virtual Machines
   - Cloud Computing
   - Big Data – Volume & Velocity

    - Re-architected High-Performance Networking – SDN & NFV
    - Building large scalable systems
3. Using Multiprocessor cluster
    - MPI
4. Multiprocessors Systems
    - With OpenCL
5. Case Studies of Large Systems
    - Case Study 1 – Large Scale Federated Systems – Mobile Communication
    - Case Study 2 – IOT (Cyber Physical System)

Lab Exercises

The first 4 or 5 exercises are for understanding system configuration, resource usage and measurement. The objective of the next few exercises are building large systems with VM and SDN. The objective of the last two exercises are developing applications with OpenCL and MPI. Case Studies are group exercise for understanding large high performance systems (understanding high performance design and deployment), no hands-on here. In general, the focus is getting familiar with terms and tools (both development, test, and measurement) with basic exposure. In-depth coverage on any of the areas is not guaranteed.

1. Introductory Material: Laptop Configuration, resource usage monitoring
    1. Understanding Laptop Configuration using Windows built-in features
        - Identify your *Persistent Storage Vendor*, Storage Type, Capacity, Maximum Sequential Read/Write Rate and Maximum Random Read/Write Rate.
        - Optionally (for Magnetic storage type) find RPM, Seek Time, Rotational Delay, Transfer Time, and any other access time related parameters. Compute the access time for 10 MB data based on this specification.
        - Find the *MTBF* value for your storage. Research and report how *MTBF* is computed by the vendor for specification.
        - Review and understand the content of the following storage specification: https://www.seagate.com/www-content/product-content/sv35-fam/sv35.5/en-us/docs/surveillance-hdd-data-sheet-ds-1679-15-1509us.pdf
        - Compute the availability value of *RAID 0, 1, 1+0, 0+1* configurations with your hard disk with availability specification found and *MTTR* of 10, then *MTTR* of 30 hrs. If your persistent storage is SSD, assume availability of 0.99.
        - Review the content of the following URL: https://www.wepc.com/tips/ssd-reliability/
        - Identify the number processors, number of cores in your processor, Vendor, model type, Operating Frequency, Threads/core. Compute the number of logical processors on your laptop.
        - Capacity of your RAM
        - Display Adapter Vendor and Adapter Model Number
        - Find out the commercial processor with largest number of cores.
        - How does the number of cores affect compiler design?

    2. Understanding multiprocessor-architecture of desktop/laptop
        - Install *Geeks3D GPU Caps Viewer* tool and find out which versions *of OpenGL, OpenCL* versions are supported
        - Play with the various tabs in the tool and get familiar with the tool.
        - We will be revisiting this tool later.

    3. Performance measurement tools of desktop/laptop with Windows built-in tools
        - Explore and find out what tools are available on Windows for resource and performance monitoring of processor, disk, and RAM usage by different processes. Learn the use of *Resource Monitor* (Control Panel → Admin Tools) on Windows.
        - Use the following documentation for understanding how to use Resource Monitor (*resmon*):
            1. https://www.digitalcitizen.life/how-use-resource-monitor-windows-7
            2. https://www.ghacks.net/2017/12/28/a-detailed-windows-resource-monitor-guide/
        - Find out how much RAM is available on the average – 1 sample per minute. Ten samples. Express the value as percentage of your RAM capacity.
        - Write a large file to disk and observe the Peak Disk Activity Rate. Check if this matches the Vendor Specification for Sequential Write.
        - How many TCP connections are opened by your browser? Record it when you are not using your browser.
        - Observe how number of Bytes/received change for your browser processor when you reload your FB Home page.
        - Now you have come this far, I can reveal this to you, there are some better tools available for Windows – use *Aida64* to do all of the above, however free version of *Aida4* is limited in feature set. It is comprehensive tool if you can afford it. More such tools listed and reviewed here https://windowsreport.com/monitor-system-resources-windows-10/
        - Find out how utilization of each core changes with no activity to intensive browsing with Aida64.
        - Find out the Cache configuration.
        - Find out the process with the largest Swap Space.
    4. Install *Wireshark* (Open Source Network monitoring tool) on Laptop. I will give you Tutorial of Wireshark. Brush up your networking basics. In a large system, network performance is a key metric. For now, an exposure to Wireshark is fine with some TCP metrics.
2. Virtual Machines: Install the *Oracle VM Virtual Box*. A short tutorial will be provided after the installation for configuration.
3. Software Thread, child Process with Client-Server Applications with VMs – Use Linux "tops" for resource usage
    - Develop a Server program that copies a file from one folder to another folder with a version number. File size should be more than 1 MB (text file is fine). Maximum of 3 *concurrent* clients are supported by this server. Each client is handled by a child process. Do the same with 4, and 5 clients (*socket* ***listen*** set to 4 and then 5).

- Measure the server performance for each client with *Server Response Time*. You need to code the time measurement in client program.
- Measure resources (CPU, RAM, and Disk) usage for different clients – snapshots with the available tools will do. Need not be process specific.
- Replace child processes with threads (*pthread*) and repeat the other steps.
- Check if there is any difference to the *overall* resource usage.

Virtualization – VM, Docker operation: Large scalable systems are built with Cloud service paradigm. The elastic capacity of cloud scales to Millions of users when needed and can scale down gracefully when those users are no longer there. Cloud paradigm is made possible by several other technologies – Virtualization, SDN, Service and Resource Orchestration etc. There is no better example of High-Performance Computing than the current data centres that host Google Gmail, Amazon eRetailing, FB Social Network, etc. The services offer high availability, elastic capacity, robust, responsive, viable, and user-friendly computing by orchestrating large volume of computing resources (CPU, Storage, RAM, and Network). Video Streaming and Mobile networking are on par with those systems. Legacy Mobile systems are migrating to Cloud Data centres by realizing the value proposition of Cloud. Virtualization allows us study many of these technologies with a single laptop!

4. Software Defined Network with Mininet

   Software Defined Network (SDN): SDN is changing the way the application

5. Cloud with OpenStack
6. Setting up Server Clusters
7. Big Data – Volume – Hadoop – Embedded/Dedicated?
8. Big Data – Velocity with Bloom filter
9. MPI Cluster with MPI
10. Open CL
11. *Optional Case Studies:*
    - *Case Study 1 – Gmail*
    - *Ca*se Study 2 – Facebook
    - Case Study 3 – Amazon
    - Case Study 4 – YouTube/Netflix
    - Case Study 5 – Mobile Communication