

# NSSA-220

## Project 1

Application Performance  
Monitoring

# Project 1 Preliminaries

- Done in teams of 3
  - Declare your team on myCourses discussion area
  - Yes, you must do this in teams! No lone wolves allowed!
- Projects are designed to be more difficult than labs and require your own research beyond what you've been taught so far (worth 15% of your grade)

# Application Performance Monitoring (APM)

- Monitoring application performance is done for many reasons including, but not limited to:
  - Detecting systemic problems with software, such as memory leaks
  - Determining if computing resources are sufficient for the mix of software being used (number of CPUs, network bandwidth, memory, disk space, disk speed)
  - Detecting malicious activity in a system

# APM Tool Overview

- You'll be writing an APM tool in Bash that will monitor a mix of processes in the form of executable programs written in C
- The script will start the processes, collect performance metrics, and perform a clean up at the end to kill these processes and any other processes that your script spawns

# APM Tool Overview (continued)

- The APM tool will monitor application performance by collecting metrics at both:
  - Process level
  - System level
- Your final submission will include a short report with Excel plots of the performance metrics collected

# Process Level Metrics

- Linux provides the ability to collect performance metrics at the process level for both CPU and memory
- Using `ps`, you can see the %CPU and %memory that any process is using
- If the applications are utilizing most of or all of the CPU, you may consider adding cores to a VM or processors to a physical machine
- Unbounded memory growth over time indicates poor memory management

# System Level Metrics

- Disclaimer: the following metrics *could* be measured at a process level, but not without great pain
- You will monitor these system level metrics
  - Network bandwidth utilization
  - Hard disk access rates
  - Hard disk utilization

# Network Bandwidth Utilization

- Network bandwidth utilization indicates how much bandwidth each network interface card is using over time
- You will measure bandwidth utilization in terms of receive (RX) data rate and transmit (TX) data rate using the ifstat tool
- If data rates are approaching network capacity, additional capacity and/or load balancing is likely needed



# Hard Disk Access Rates

- Hard disk access rates are typically measured in volume of reads and writes per second
- Monitoring disk access activity over time may show that a solid state drive is needed over a mechanical drive or alert a system administrator to unexpected excessive access rates
- You will measure hard disk writes in kB/second to the primary hard drive (sda) using the iostat tool

# Hard Disk Utilization

- Monitoring hard disk utilization, in terms of remaining disk capacity, is not as critical as it once was, but there are still times where unexpected *reductions* or *increases* in disk capacity may occur
- You will measure hard disk utilization on the “/” mount (the /dev/mapper/centos-root filesystem) using the df tool

# CentOS VM Resources

- The default CentOS VM resources are
  - 1 CPU core
  - 2 GB memory
  - 100 GB hard disk
  - CentOS NIC (ens33) in NAT mode
- You are required to use the default CentOS VM resources to collect data for this project

# Applications to Monitor

- You will be given 6 applications in the form of pre-compiled C executables
- Each application will take a single argument: the IP address of a VMNet NIC on the lab PCs or any virtual NIC on your own computer
- You will need to run your VM in “NAT Mode” to allow connectivity between the VM and the host’s virtual NIC. The virtual NIC is initially disabled.

# APM Tool Requirements

1. The APM tool shall collect process and system level metrics every 5 seconds
2. The APM tool shall collect %CPU and %memory utilization *per process* using the ps tool
3. The APM tool shall collect network bandwidth utilization in terms of RX data rate and TX data rate (kB/s) with a *sampling interval of 1 second* on the ens33 interface using the ifstat tool

# More APM Tool Requirements

4. The APM tool shall collect hard disk writes in kB/second to the primary hard drive (sda) using the iostat tool
5. The APM tool shall collect hard disk utilization of the “/” mount in *Megabytes available* using the df tool
6. The APM tool shall output all CPU and memory metrics to a CSV file specific to the process they were measured from. Name the files `<proc_name>_metrics.csv`

# More APM Tool Requirements

7. The format of the process specific output files shall be
  - <seconds>, <%CPU>, <%memory>
8. The APM tool shall write all system level metrics to a file called `system_metrics.csv`
9. The format of the system level output file shall be
  - <seconds>, <RX data rate>, <TX data rate>, <disk writes>, <available disk capacity>

# More APM Tool Requirements

- 10. The APM tool shall spawn all application processes
- 11. The APM tool shall kill all application processes and any other processes it spawns in an exit trap function called “cleanup”
- 12. The APM tool shall minimally include functions to (1) spawn applications and other processes, (2) collect process level metrics, and (3) collect system level metrics



# More APM Tool Requirements

13. Your team's submission to myCourses shall be a single zip file including the APM tool script, all output files for a *15 minute run*, and a report showing Excel plots of the metrics collected. The report format is provided.
14. Your application performance data shall be measured using the default VMWare resources

# Grading

- See grading sheet for details
- Major point loss will occur if an individual exhibits poor teamwork. You will fill out a peer review form to indicate how well the team worked together. Let me know ASAP if there are issues with your team.

# Hints

- Read the man pages for the specified tools!
- One of the tools requires you to spawn an additional process to use it.
- All the values in the output files should be numbers only. You already know what the units are.
- Put all the requirements into a table in a Word document and check them off as you complete them.

# Points of Inspiration

- APM is highly relevant to just about any career path in computing, especially systems engineering, network management, network and system security, and system administration
- Employers are impressed by anyone that has experience with APM and analyzing APM data is highly related to the field of data analytics/data science

# Ask for help!

- Don't suffer in silence. Ask me or your TA for help sooner rather than later!
  - Attend my office hours or the TA's
  - Make an appointment outside of office hours
  - Send an email
- If you're not sure if you've met a specific requirement, please ask!