2024 Summer Research Training
Tasks for Cloud Technologies Teams / Week 1

Announced: 14.08.2024
Deadline: 21.08.2024 23.59

Create a github repo named SRT2024-LLM-Team-X (X is your team number).
Submit all your work to your github repo. Keep the repo organized. Add at-ay to your repo as a collaborator.

→ You will have both a theoretical and a practical work for the Week 1.

Theoretical Work

Aim: Understand how basic cloud technologies work. Carry out detailed research as a group. Frequently discuss your findings between members. At the end you will prepare a 7 mins presentation

Some questions for guiding your theoretical work (You are not limited with the below outline)

What is virtualization, and how does it contribute to efficient resource management in cloud environments?

How do hypervisors work, and what are the differences between Type 1 and Type 2 hypervisors?

What are the advantages and disadvantages of using virtual machines compared to physical servers?

Explain the concept of containers and how they differ from virtual machines.

Why are containers considered more lightweight than virtual machines?

What kind of impact do containers have on software deployment and development?

What role does container orchestration play in modern cloud infrastructure?

How does Kubernetes manage containerized applications at scale? Discuss its core components. (include kubernetes architecture diagram and make sure you are able to discuss how the components work)

Practical Work (Every member of the group should do the following steps)

Part I (VM)
→ Use a tool like Virtualbox or Vmware to create a VM on your local machine. Install Ubuntu LTS server on this VM.
→ Deploy a web server on this VM
→ Use a tool like ab (apache benchmark) to stress test your system. Report how the response time changes when you increase the load.

Part II (Kubernetes)
→ Deploy Kubernetes on your local machine. (First create necessary VMs on your local machine for this process)
→ Deploy a web application on your Kubernetes cluster. Make sure you can access this application from your local machine.
→ Use a tool like ab (apache benchmark) to stress test your system. Report how the response time changes when you increase the load.
(Optional) Turn on auto-scaling options on Kubernetes and stress test your application again. Report how the response time changes when you increase the load.