Group Assignment 5: Paper Review
CECS 326 –Operating Systems

1. Summary

This assignment requires to thoroughly review one high-quality research paper from famous conferences or journals. You should select one interested paper from the given paper list and summarize your review into a file.

You should submit the required deliverable materials on Canvas by **11:55pm, May 12th (Sunday), 2024**.

2. Description

Reading a good research paper can greatly improve your perspective and interests for one particular research area. Here list several papers that are from high qualified or famous conferences, and each represents one excellent insight or improvement in a particular real problem. Please only **select one paper** from the below list:

- Help Rather Than Recycle: Alleviating Cold Startup in Serverless Computing Through Inter-Function Container Sharing, ATC 2022

- Faith: An Efficient Framework for Transformer Verification on GPUs, ATC 2022

- Meces: Latency-efficient Rescaling via Prioritized State Migration for Stateful Distributed Stream Processing Systems, ATC'22

- Towards Latency Awareness for Content Delivery Network Caching, ATC'22

- Investigating Managed Language Runtime Performance: Why JavaScript and Python are 8x and 29x slower than C++, yet Java and Go can be Faster? , ATC'22

- SecretFlow-SPU: A Performant and User-Friendly Framework for Privacy-Preserving Machine Learning, ATC'23

- Portunus: Re-imagining Access Control in Distributed Systems, ATC'23

- TC-GNN: Bridging Sparse GNN Computation and Dense Tensor Cores on GPUs, ATC'23

- Comosum: An Extensible, Reconfigurable, and Fault-Tolerant IoT Platform for Digital Agriculture, ATC'23

- Oakestra: A Lightweight Hierarchical Orchestration Framework for Edge Computing, ATC'23

- Explore Data Placement Algorithm for Balanced Recovery Load Distribution, ATC'23

- Sponge: Fast Reactive Scaling for Stream Processing with Serverless Frameworks, ATC'23

- Decentralized and Stateful Serverless Computing on the Internet Computer Blockchain, ATC'23

- AWARE: Automate Workload Autoscaling with Reinforcement Learning in Production Cloud

Systems, ATC'23

- Nodens: Enabling Resource Efficient and Fast QoS Recovery of Dynamic Microservice Applications in Datacenters, ATC'23
- Tectonic-Shift: A Composite Storage Fabric for Large-Scale ML Training, ATC'23
- EnvPipe: Performance-preserving DNN Training Framework for Saving Energy, ATC'23
- UnFaaSener: Latency and Cost Aware Offloading of Functions from Serverless Platforms, ATC'23
- Accelerating Distributed MoE Training and Inference with Lina, ATC'23
- MSRL: Distributed Reinforcement Learning with Dataflow Fragments, ATC'23
- Nimble: Rollback Protection for Confidential Cloud Services, OSDI 23
- An Extensible Orchestration and Protection Framework for Confidential Cloud Computing, OSDI 23
- Hyrax: Fail-in-Place Server Operation in Cloud Platforms, OSDI 23
- Defcon: Preventing Overload with Graceful Feature Degradation, OSDI 23
- AlpaServe: Statistical Multiplexing with Model Parallelism for Deep Learning Serving, OSDI 23
- Cocktailer: Analyzing and Optimizing Dynamic Control Flow in Deep Learning, OSDI 23
- Automated Verification of Idempotence for Stateful Serverless Applications, OSDI 23

How to Get a Copy of Paper:

Once you've found the citation for a paper that is relevant to your advanced science project, the next step is getting a copy so that you can read it. Some search engines provide links to free online versions of the paper, if one exists. There are several ways to find the paper or copies:

- **Look for a free online version.** Try searching for the full title of the paper in a regular search engine like Google Scholar, Yahoo, or Microsoft Scholar. The paper may come up multiple times, and one of those might be a free, downloadable copy. So, if the first link isn't downloadable, try another. Tips: For *Google Scholar*, click the "All # versions", you may find some versions can be freely accessed.

- **Search directly for the homepage of the first or last author** of the paper and see if he or she has a PDF of the paper on his or her website. If so, you can download it directly from there. Generally, it is only worth looking up the first author (the one who contributed the most to the paper) or the last author (usually the professor in whose lab the work was done and who supervised the science project).

- **Go directly to the online homepage of the journal** in which the paper was published. Some scientific journals are "open-source," meaning that their content is always free online to the public. Others are free online (often after registering with the website) if the paper

- **Utilize the CSULB network.** You can utilize the "CSULB Library" (the link is in the above table) to access the articles with your student account.

Reading Instructions:

Review the paper should concentrate on the motivation and the design methodology of the proposed work. You should pay attention to the author's motivation, design novelty, proposed system/architecture, and comparison with previous work. You may skip the technical details that you are not familiar with. You should focus on the improvement or novel part of the proposed work.

2. Write Instructions

You should write the summary via **Word** or **Latex**. The submitted file **must be a pdf format**.

Component 1: Summary of paper

In this component, you should use one or two paragraphs to summarize the paper your read. You should introduce the problem they provided, the design motivation, design novelty, proposed architecture/methodology.

Component 2: Your views or analysis

In this component, you should provide your analysis or perspectives about the paper. You should comment on their proposed problem, discuss their design, and provide your own views about their work (try to catch up with the fundamental knowledge from OS courses). For example, you can point out the possible improvement of the proposed solution, or clarify the drawback or inefficiency of their work from any point, or the differences from the topics that you had learned from class. You can also analyze the difference from previous work, and show your perspective of their novelty, or classify the possible shortcoming in their design. This component will be the majority part of your report.

Formats requirements:

- We will do similarity check. The accepted similarity must be **less than 20%**, otherwise it will be taken as plagiarism!!
- **NO AI tools** for the assignment (ChatGPT, GPT4, etc). We will run anti-AI tools to detect it. If more than 20% of contents are labelled as generated from AI, then it will be taken as plagiarism!!
- The length of your report is around *1.5 pages*. (Except the title and name info, etc.)
- The font should be Times New Roman, Size 11, single line space, 0 bt in before or after space of paragraph, 8.5*11'.
- The submission only includes *one pdf file*, no other files should be submitted.
- Group members should both contribute to the assignment, 0 grade if has no contributions to the submitted assignment.

3. Grades Criteria:

Details	Points
Submitted file with correct format.	20 pts
Detailed paper summary.	20 pts
Details of your analysis or discussion.	60 pts