2025 Spring Systems Reading Group

Welcome Everyone!

Jiyang Wang & Kunzhao Xu

2025.02.25

Agenda

- **Introduction to Reading Group**
  - Mission
  - Arrangement
  - Format & Requirements

- Advices for reading a paper

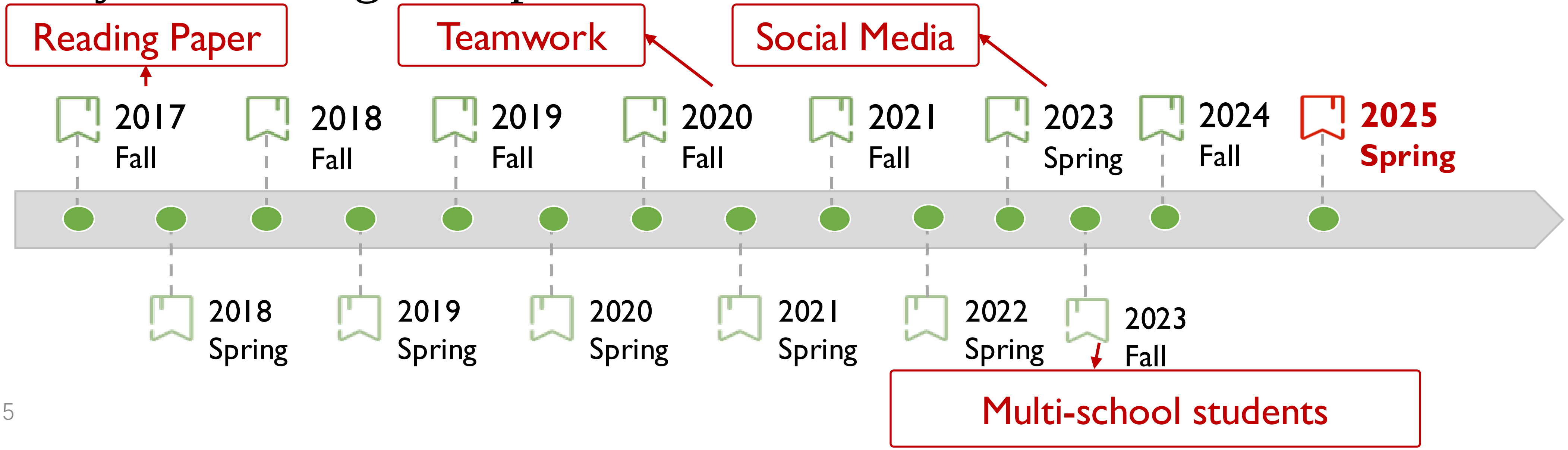
- Advices for giving a talk

Mission of reading group

- Understand and keep abreast of “latest research in **systems research**”
- Learn “how to do **high-quality** systems research”
- Polish soft skills
  - Understanding
  - Presentation
  - Critical thinking
  - Communication
  - …

Mission of reading group

- Understand and keep abreast of “latest research in **systems research**”
- Learn “how to do **high-quality** systems research”
- History of Reading Group



Reading Paper

Teamwork

Social Media

2017 Fall

2018 Spring

2018 Fall

2019 Spring

2019 Fall

2020 Spring

2020 Fall

2021 Spring

2021 Fall

2022 Spring

2023 Spring

2023 Fall

2024 Fall

2025 Spring

Multi-school students

Mission of reading group

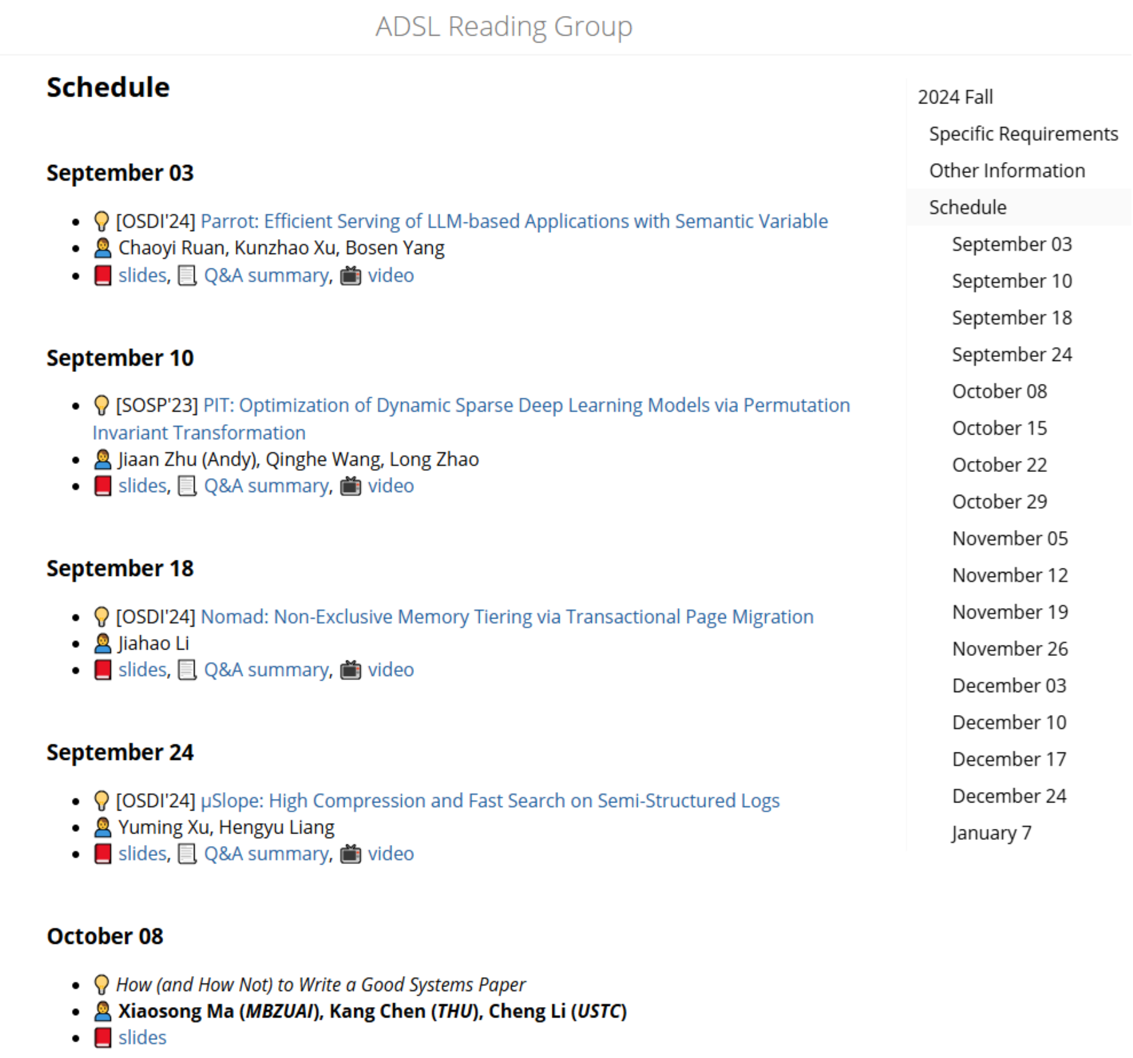
- Understand and keep abreast of “latest research in **systems research**”

- Learn “how to do **high-quality** systems research”

- Target of this semester
  - **Paper Sharing**
    - Improve the presentation quality
    - More discussion and brainstorming
  - **More than one paper**
    - Choose one more paper from arXiv

Previous RG

- We read papers from:
  - SOSP’ 23, 24
  - OSDI’ 24
- 17 presentations were given
- Presenters were from
  - USTC ADSL
  - Tianjin University
  - Northwestern Polytechnical University
  - …



ADSL Reading Group

Schedule

September 03

- 💡 [OSDI'24] Parrot: Efficient Serving of LLM-based Applications with Semantic Variable
- 👤 Chaoyi Ruan, Kunzhao Xu, Bosen Yang
- 📕 slides, 📄 Q&A summary, 🎬 video

September 10

- 💡 [SOSP'23] PIT: Optimization of Dynamic Sparse Deep Learning Models via Permutation Invariant Transformation
- 👤 Jiaan Zhu (Andy), Qinghe Wang, Long Zhao
- 📕 slides, 📄 Q&A summary, 🎬 video

September 18

- 💡 [OSDI'24] Nomad: Non-Exclusive Memory Tiering via Transactional Page Migration
- 👤 Jiahao Li
- 📕 slides, 📄 Q&A summary, 🎬 video

September 24

- 💡 [OSDI'24] μSlope: High Compression and Fast Search on Semi-Structured Logs
- 👤 Yuming Xu, Hengyu Liang
- 📕 slides, 📄 Q&A summary, 🎬 video

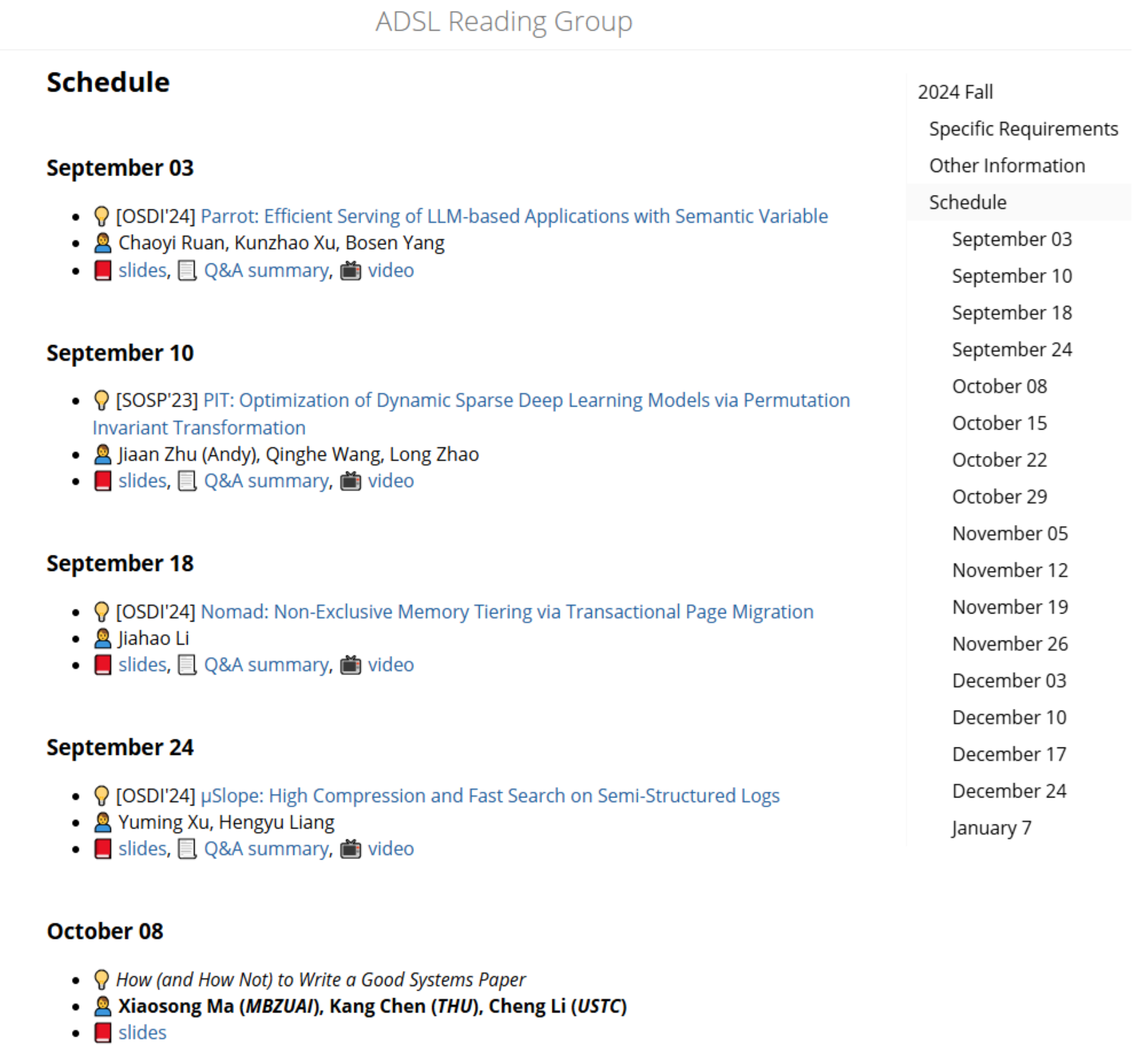
October 08

- 💡 *How (and How Not) to Write a Good Systems Paper*
- 👤 **Xiaosong Ma (*MBZUAI*), Kang Chen (*THU*), Cheng Li (*USTC*)**
- 📕 slides

Previous RG

- Topic
  - Storage / Memory
    - Page migration
    - CPU Stall
    - Disaggregated memory
    - ZNS-SSD
  - LLM / AI
    - Latency optimization
    - Serverless
    - KV Cache
    - Parallelism
  - How to Write a Good Systems Paper

……



ADSL Reading Group

Schedule

September 03

- 💡 [OSDI'24] Parrot: Efficient Serving of LLM-based Applications with Semantic Variable
- 🧑 Chaoyi Ruan, Kunzhao Xu, Bosen Yang
- 📕 slides, 📄 Q&A summary, 📺 video

September 10

- 💡 [SOSP'23] PIT: Optimization of Dynamic Sparse Deep Learning Models via Permutation Invariant Transformation
- 🧑 Jiaan Zhu (Andy), Qinghe Wang, Long Zhao
- 📕 slides, 📄 Q&A summary, 📺 video

September 18

- 💡 [OSDI'24] Nomad: Non-Exclusive Memory Tiering via Transactional Page Migration
- 🧑 Jiahao Li
- 📕 slides, 📄 Q&A summary, 📺 video

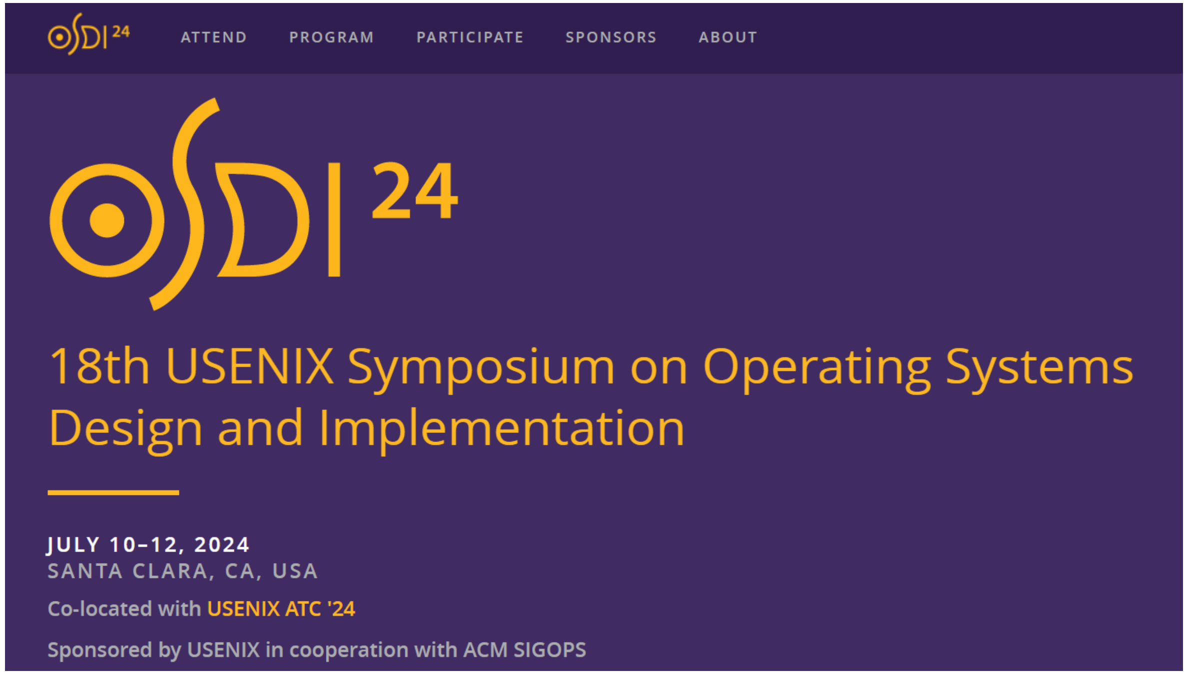
September 24

- 💡 [OSDI'24] μSlope: High Compression and Fast Search on Semi-Structured Logs
- 🧑 Yuming Xu, Hengyu Liang
- 📕 slides, 📄 Q&A summary, 📺 video

October 08

- 💡 *How (and How Not) to Write a Good Systems Paper*
- 🧑 **Xiaosong Ma (*MBZUAI*), Kang Chen (*THU*), Cheng Li (*USTC*)**
- 📕 slides

What do we read?

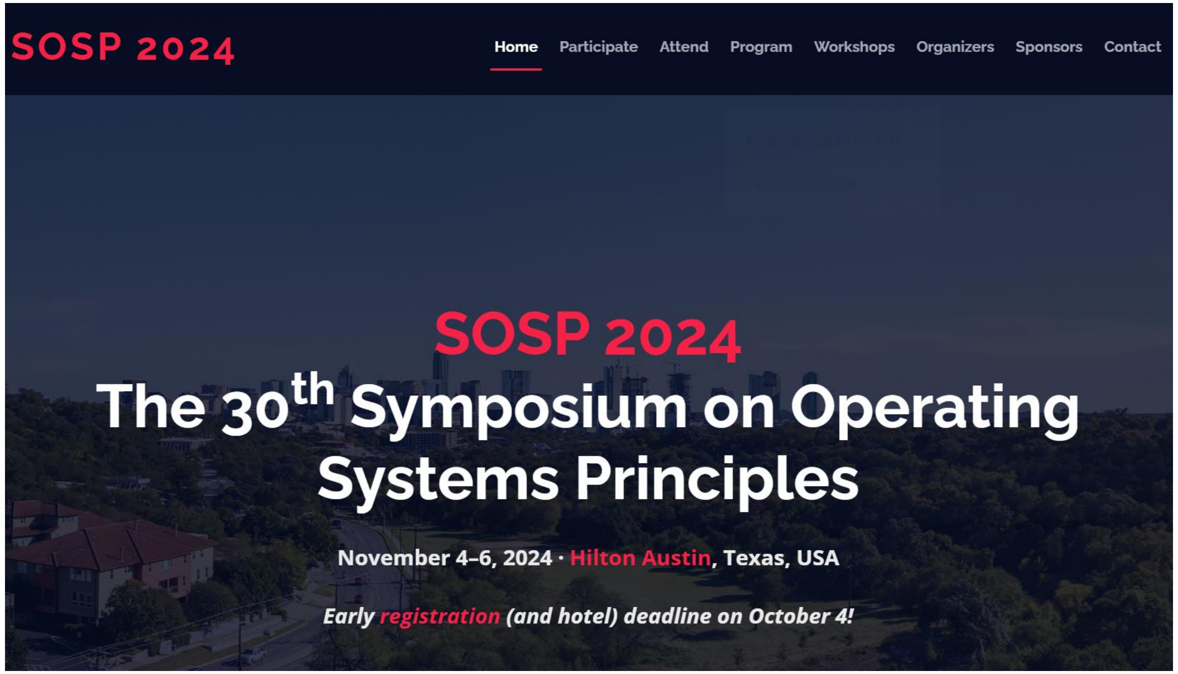


18th USENIX Symposium on Operating Systems Design and Implementation

JULY 10–12, 2024
SANTA CLARA, CA, USA

Co-located with USENIX ATC '24

Sponsored by USENIX in cooperation with ACM SIGOPS



SOSP 2024

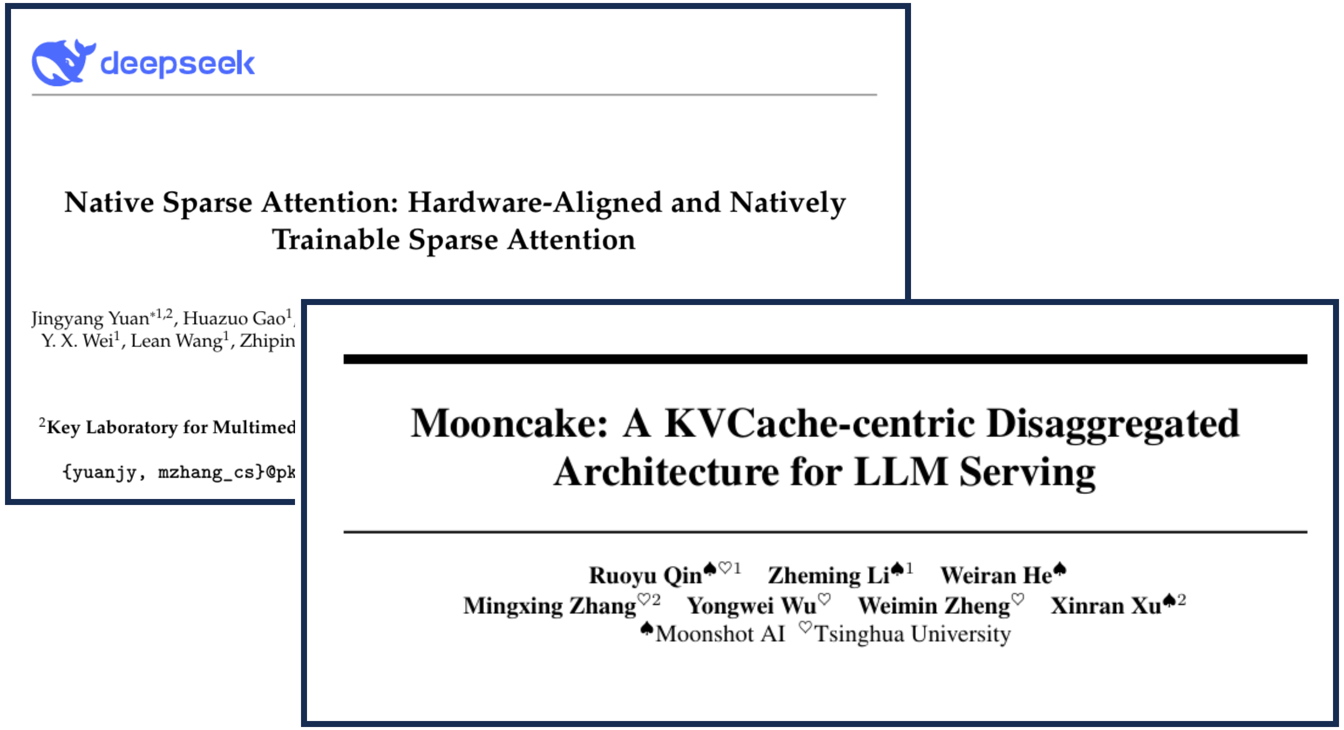
The 30th Symposium on Operating Systems Principles

November 4-6, 2024 · Hilton Austin, Texas, USA

Early registration (and hotel) deadline on October 4!

Read best papers!!!

What do we read?



deepseek

Native Sparse Attention: Hardware-Aligned and Natively Trainable Sparse Attention

Jingyang Yuan[*1,2], Huazuo Gao[1],
Y. X. Wei[1], Lean Wang[1], Zhipin

[2]Key Laboratory for Multimed

{yuanjy, mzhang_cs}@pk

Mooncake: A KVCache-centric Disaggregated Architecture for LLM Serving

Ruoyu Qin[♠♡1] Zheming Li[♠1] Weiran He[♠]
Mingxing Zhang[♡2] Yongwei Wu[♡] Weimin Zheng[♡] Xinran Xu[♠2]
[♠]Moonshot AI [♡]Tsinghua University

⚠ **Read latest papers!!!**

Paper sharing: arrangement

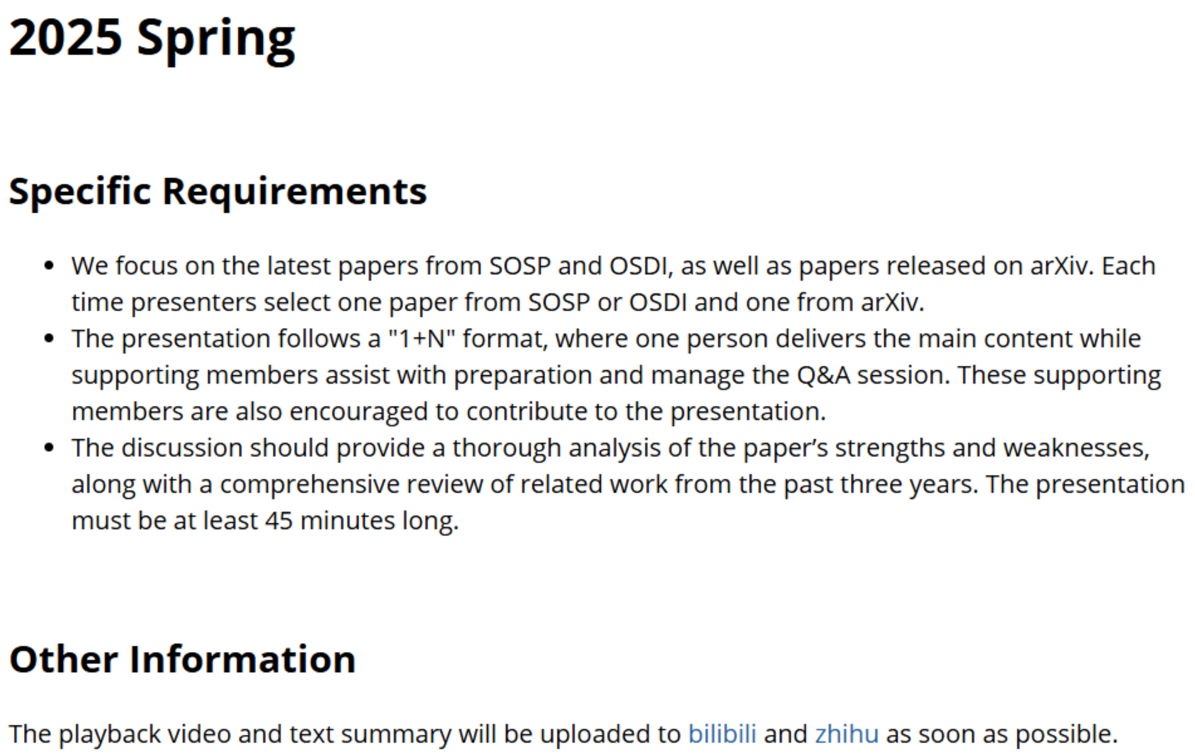
- Time: 19:00 – 21:00, every Tuesday

- Location:
  - Offline: 高新区信智楼A707
  - Online: Tencent meeting 877-6724-4752

- Webpage: https://adsl-rg.github.io/2025_spring.html

Paper sharing: arrangement

- Time: 19:00 – 21:00, every Tuesday
- Location:
  - Offline: 高新区信智
  - Online: Tencent meet
- Webpage: https://adsl-rg.



2025 Spring

Specific Requirements

- We focus on the latest papers from SOSP and OSDI, as well as papers released on arXiv. Each time presenters select one paper from SOSP or OSDI and one from arXiv.
- The presentation follows a "1+N" format, where one person delivers the main content while supporting members assist with preparation and manage the Q&A session. These supporting members are also encouraged to contribute to the presentation.
- The discussion should provide a thorough analysis of the paper’s strengths and weaknesses, along with a comprehensive review of related work from the past three years. The presentation must be at least 45 minutes long.

Other Information

The playback video and text summary will be uploaded to bilibili and zhihu as soon as possible.

Paper sharing: arrangement

- Each presentation led by two students
  - Choose the paper (one paper from OSDI or SOSP and one from arXiv)
  - Find your teammates (one team for OSDI/SOSP paper and the other for arXiv)
  - **Guarantee the quality**
  - Presentation video: Upload to bilibili
- We also encourage students from other schools or labs to participate in the RG :)

Paper sharing: format

- Primary focus: **understanding the paper**
  - What is the problem?
  - What are the challenges?
  - What are state-of-the-arts, and their deficiencies?
  - What are the key insights/techniques?
  - Lessons learned from experiments?
- Whole discussion: 1.5~2 hours, presentation: **70~80 minutes**

Paper sharing: tips

- Please make around **70 slides**!
  - Too much text ☹
  - Copy paste figures ☹
  - Animations ☺
  - Transitions between slides ☺

- One slide: 1 - 2 minutes

- Please do rehearsals offline

Paper sharing: tips

- Please make around **70 slides**!
  - Too much text ☹
  - Copy paste figures ☹
  - Animations ☺
  - Transitions between slides ☺

- One slide: 1 - 2 minutes

- Please do rehearsals offline

- Additional requirement:
  - **A mind map**
  - **Summary after sharing**
    - Problem
    - Key insights/techniques
    - Evaluation
    - Strengths
    - Improvement
    - Record Q&A (by Jiyang & Kunzhao)
    - Submit to 知
      (by Jiyang & Kunzhao)

Ready to share?

- Please make around **70 slides**!
  - Too much text ☹
  - Copy paste figures ☹
  - Animations ☺
  - Transitions between slides ☺

- Additional requirement:
  - **A mind map**
  - **Summary after sharing**
    - Problem
    - Key insights/techniques
    - Evaluation

Ready to share? Fill the follow document!

https://docs.qq.com/sheet/DRWdyZVpGTIJKSWJR

If you are from other schools or labs, let us know :)

Agenda

- Introduction to Reading Group
  - Mission
  - Arrangement
  - Format & Requirements

- **Advices for reading a paper**

- Advices for giving a talk

How to read a paper!

- From Srinivasan Keshav
  - The Robert Sansom Professor of Computer Science at the University of Cambridge
  - ACM/IEEE Fellow

- **Three passes**
  - 1st: get a bird's-eye view
  - 2nd: grasp the content
  - 3rd: rethink, recreate the work

- http://ccr.sigcomm.org/online/files/p83-keshavA.pdf

Agenda

- Introduction to Reading Group
  - Mission
  - Arrangement
  - Format & Requirements
- Advices for reading a paper
- **Advices for giving a talk**

Advices

- https://people.eecs.berkeley.edu/~jrs/speaking.html
  - Preparing a talk
  - Giving the talk
- http://pages.cs.wisc.edu/~markhill/conference-talk.html
  - Oral presentation advice
  - How to give a bad talk

2025 Spring
Systems Reading Group

Q& A

USTC, CHINA

ADSLAB

先进数据系统实验室