CS341 - Advanced Database Topics @ KAUST

VENUE: B1-4320 (CS lab - fishbowl room at the middle of level 4)
TIME: Sunday, 9:30am-11:00am and Wednesday, 9:30am-11:00am

Pre-requisites: Student must have taken CS245 or equivalent course and must have excellent programming experience, preferably in C/C++ and Linux.

Topics: The course will focus on Data Management on Parallel and Distributed systems. Topics include: Distributed joins, Distributed Hash Tables, CAN, Chord, the Map-Reduce framework, Eddies, implementation of relational operators on multicore architectures, databases in the Cloud and others. Particular attention will be given to recent research on parallel and distributed systems for Machine Learning and AI.

- Every lecture will focus on one (or at most 2) research paper(s). All students will have to read the paper and write a short review (roughly 150 words).
- One student will present the paper (roughly 30-40min) and then the whole class will discuss the pros and cons of the work.
- Each student must complete a substantial programming project. This year's topic will be the implementation of an algorithm/system related to large graphs. At the very least, the system must run on one GPU; hopefully, it will run on a large number of GPUs. Based on these guidelines, each student must **define their own project**.

Reading material: The course will be based on research papers. All papers can be found on the course's wiki: http://cs341.pbworks.com. Check the wiki to request access. No textbook is needed. The library has several books on databases that can be used for reference.

Assessment:

- Programming project: 40% (deadline: 11-May-2023)
- Oral presentations: 20%
- Paper reviews: 20%
- Class participation: 20%

There will be NO final exam

Lecturer: Panos Kalnis http://web.kaust.edu.sa/faculty/PanosKalnis

Office hours: I am available almost every day. Send me an email (panos.kalnis@kaust.edu.sa) to schedule an appointment. Office: B1-4416 (at the back of the building)

Papers

The schedule below is tentative. It may be rearranged, depending on seminars by visitors. Papers may be added or removed, based on the background of the students and the requirements of the projects.

- MiCS: Near-linear Scaling for Training Gigantic Model on Public Cloud
Zhen Zhang, Shuai Zheng, Yida Wang, Justin Chiu, George Karypis, Trishul A Chilimbi, Mu Li, Xin Jin
https://www.vldb.org/pvldb/vol16/p37-zhang.pdf

- DataMUX: Data Multiplexing for Neural Networks
https://arxiv.org/abs/2202.09318

- Harmony: Overcoming the Hurdles of GPU Memory Capacity to Train Massive DNN Models on Commodity Servers
https://www.vldb.org/pvldb/vol15/p2747-li.pdf

- VARUNA - EuroSyS 2022

- DBOS: A DBMS-oriented Operating System
Athinagoras Skiadopoulos, Qian Li, Peter Kraft, Kostis Kaffes, Daniel Hong, Shana Mathew, David Bestor, Michael Cafarella, Vijay Gadepally, Goetz Graefe, Jeremy Kepner, Christos Kozyrakis, Tim Kraska, Michael Stonebraker, Lalith Suresh, Matei Zaharia, http://www.vldb.org/pvldb/vol15/p21-skiadopoulos.pdf

ZeRO: Memory Optimizations Toward Training Trillion Parameter Models, Samyam Rajbhandari, Jeff Rasley, Olatunji Ruwase, and Yuxiong He, SC, 2020.

Near-Optimal Sparse Allreduce for Distributed Deep Learning, Shigang Li, Torsten Hoefler, Proceedings of the 27th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP'22), 2022.

p00. Introduction: The long game of research, MY Vardi, CACM, 2019.

p01. The Gamma Database Machine Project, D.J. DeWitt, S. Ghandeharizadeh, D.A. Schneider, A. Bricker, H. Hsiao, R. Rasmussen, *IEEE TKDE, 2(1)*, pp. 44-62, 1990.
p02. Mariposa: a wide-area distributed database system, M. Stonebraker, P.M. Aoki, W. Litwin, A. Pfeffer, A. Sah, J. Sidell, C. Staelin, A. Yu, *The VLDB Journal, 5(1)*, pp. 48-63, 1996.

p03. Chord: a scalable peer-to-peer lookup protocol for internet applications, I. Stoica, R. Morris, D. Liben-Nowell, D.R. Karger, M.F. Kaashoek, F. Dabek, H. Balakrishnan, *IEEE/ACM Transactions on Networks*, 11(1), pp. 17-32, 2003.

p04. The case for RAMClouds: scalable high-performance storage entirely in DRAM, John Ousterhout, et al, SIGOPS Oper. Syst. Rev. 43, pp. 92–105, 2010.

p05. The Google File System. S. Ghemawat, H. Gobioff, S.T. Leung, In Proc. of ACM Symposium on Operating Systems Principles (SOSP),pp. 29-43, 2003.

p06. MapReduce: Simplified Data Processing on Large Clusters. J Dean and S. Ghemawat, Proc. of Symposium on Operating System Design and Implementation (OSDI), 2004

p07. Bigtable: A distributed storage system for structured data, F. Chang, J. Dean, S. Ghemawat, W.C. Hsieh, D.A. Wallach, M. Burrows, T. Chandra, A. Fikes, R.E. Gruber, *In Proc. of USENIX-OSDI*, pp. 205-218, 2006.

p08. Spark: Cluster Computing with Working Sets. Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, Ion Stoica. *In Proc. of HotCloud,* 2010.

p09. C-Store: A Column-oriented DBMS, M. Stonebraker, D.J. Abadi, A. Batkin, X. Chen, M. Cherniack, M. Ferreira, E. Lau, A. Lin, S. Madden, E. O'Neil, P O'Neil, A. Rasin, N. Tran, S. Zdonik, *In Proc. of VLDB*, pp. 553-564, 2005.

p10. Pregel: a system for large-scale graph processing, G. Malewicz, M.H. Austern, A.J.C Bik, J.C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski. *In Proc. of SIGMOD*, 2010.

p11. Matrix Algebra Framework for Portable, Scalable and Efficient Query Engines for RDF Graphs, Fuad Jamour, Ibrahim Abdelaziz, Yuanzhao Chen, and Panos Kalnis. In Proc. of EuroSys, 2019.

p12. Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis, Tal Ben-Nun, Torsten Hoefler, arXiv.org, 2018

p13. Exploring Autoencoder-based Error-bounded Compression for Scientific Data, Jinyang Liu, Sheng Di, Kai Zhao, Sian Jin, Dingwen Tao, Xin Liang, Zizhong Chen, Franck Cappello, arXiv.org, 2021.

p14. Project Adam: building an efficient and scalable deep learning training system. Trishul Chilimbi, et al., In Proc of OSDI, pp. 571-582, 2014.

p15. TensorFlow: A system for large-scale machine learning. M. Abadi et al., In Proc. of 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2016.

p16. Parameter Hub: a Rack-Scale Parameter Server for Distributed Deep Neural Network Training. Liang Luo et al, ACM Symposium on Cloud Computing (SoCC), 2018.
p17. Scaling Distributed Machine Learning with In-Network Aggregation, Amedeo Sapio, Marco Canini, Chen-Yu Ho, Jacob Nelson, Panos Kalnis, Changhoon Kim, Arvind Krishnamurthy, Masoud Moshref, Dan R. K. Ports, Peter Richtárik, arXiv, 2019.
p18. Compressed Communication for Distributed Deep Learning: Survey and Quantitative Evaluation, H. Xu, CY. Ho, A. Abdelmoniem, A. Dutta, EH Bergou, K. Karatsenidis, M. Canini, P. Kalnis, KAUST Technical Report, 2020.

p19. HOGWILD!: a lock-free approach to parallelizing stochastic gradient descent. Feng Niu et al., In Proc. of NIPS, ,pp. 693-701, 2011

p20. A domain-specific architecture for deep neural networks. N.P. Jouppi, C. Young, N., D. Patterson. In Commun. ACM 61(9), pp. 50-59, 2018. (August 2018), 50-59.

p21. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity, William Fedus, Barret Zoph, Noam Shazeer, arXiv, 2021.

p22. Scientific Data Compression - Tutorial by F. Capello @ SC21
Watch all videos and prepare a 500 word summary (no pros/cons)
Slides: https://drive.google.com/drive/folders/1Rd4dL3tdc1dCL1-X4ul1hhi4W8xW1Wkr?usp=sharing
Videos: https://www.youtube.com/watch?v=pm7RcbPA0tM&list=PL8c70A0sMgXm9yRVrmFAbXCblnNZc43-B

p23. Optimizing Lossy Compression with Adjacent Snapshots for N-body Simulation Data, Sihuan Li, Sheng Di, Xin Liang, Zizhong Chen and Franck Cappello, IEEE Big Data, 2018.

p24. Kinematic self-replication in reconfigurable organisms, Sam Kriegman et al., PNAS, 2021.