R244: Paper Review Presentation Assignment

2024/10/23 Session 2: Data flow programming (16 mins presentation)

wz337 (Andy) 5. Derek Murray, Malte Schwarzkopf, Christopher Smowton, Steven Smith, Anil Madhavapeddy and Steven Hand: Ciel: a universal execution engine for distributed data-flow computing, NSDI 2011.

lc869 (Luca) 6.2. D. Murray, F. McSherry, R. Isaacs, M. Isard, et al.: Naiad: A Timely Dataflow System, SOSP, 2013. **You can also read 6.1 and 6.3 if you can.**

gm799 (Gabriel) 10. Y. Yu et al.: Dynamic Control Flow in Large-Scale Machine Learning, EuroSys, 2017. **You can also read 8 and 8.2 if you can.**

ct678 (Chris) 11. R. Nishihara, P. Moritz, et al.: Ray:A Distributed Framework for Emerging AI Applications, OSDI, 2018.

as3623 (Andrzej) 16. P. Barham, et al.: Pathways: Asynchronous Distributed Dataflow for ML, MLSys, 2022.

hxl23 (Mark) 18. H. Zhu, et al.: MSRL: Distributed Reinforcement Learning with Dataflow Fragments, USENIX_ATC, 2023.

2024/10/30 Session 3: Large-scale graph data processing (max 14 mins presentation)

hs899(Hetong) 1. G. Malewicz, M. Austern, A. Bik, J. Dehnert, I. Horn, N. Leiser, and G. Czajkowski: Pregel: A System for Large-Scale Graph Processing, SIGMOD, 2010.

edjg2 (Edmund) 4. J. Gonzalez, Y. Low, H. Gu, D. Bickson, and C. Guestrin: Powergraph: distributed graph-parallel computation on natural graphs. OSDI, 2012.

oea27 (Timi) 5. J. Shun and G. Blelloch: Ligra: A Lightweight Graph Processing Framework for Shared Memory, PPoPP, 2013.

ams345 (Aryan) 9. A. Roy, I. Mihailovic, W. Zwaenepoel: X-Stream: Edge-Centric Graph Processing using Streaming Partitions, SOSP, 2013.

jkb55 (Kuba) 11. F. McSherry, M. Isard and D. Murray: Scalability! But at what COST? , HOTOS, 2015.

at2124 (Asbjørn)19. Z. Jia, Y. Kwon, G. Shipman, P. McCormick, M. Erez, A. Aiken: A Distributed Multi-GPU System for Fast Graph Processing, VLDB, 2018.

sn666 (Sid) 23. T. Kipf, et al.: Semi-Supervised Classification with Graph Convolutional Networks. ICLR, 2017.

2024/11/13 Session 5: Probabilistic Programming (max 14 mins presentation)

zc344 (Xavier) 7. V. Dalibard, M. Schaarschmidt, and E. Yoneki: BOAT: Building Auto-Tuners with Structured Bayesian Optimization, WWW, 2017.

ak2599 (Andrew) 14. J. Shao, et al.: Tensor Program Optimization with Probabilistic Programs, ArXiv, 2022.

lc869 (Luca) 20. Z. Wang, C. Li, S. Jegelka, and P. Kohli: Batched High-dimensional Bayesian Optimization via Structural Kernel Learning, PLMR, 2017.

ct678 (Chris) 30. O. Alipourfard et al.: CherryPick: Adaptively Unearthing the Best Cloud Configurations for Big Data Analytics, NSDI, 2017.

hxl23 (Mark) 1.1. A. Mirhoseini et al.: Device Placement Optimization with Reinforcement Learning, ICML, 2017.
…plus if you can read **1.2.** A. Mirhoseini, A. Goldie et al.: A Hierarchical Mode for Device Placement, ICLR, 2018.

at2124 (Asbjørn) 3. R. Addanki, S. B. Venkatakrishnan, S. Gupta, H. Mao, M. Alizadeh: Placeto: Learning Generalizable Device Placement Algorithms for Distributed Machine Learning, arXiv, 2019.

ams345 (Aryan)12. D. Aken , D. Yang, S. Brillard, A. Fiorino, B. Zhang, C. Bilien, and A. Pavlo: An Inquiry into Machine Learning-based Automatic Configuration Tuning Services on Real-World Database Management Systems, VLDB, 2021.
This paper is based on 10 Automatic database management system tuning through large-scale machine learning.

edjg2 (Edmund) 19. D. Ha and J. Schmidhuber: World Models, arXiv, 2018 (https://worldmodels.github.io).

as3623 (Andrzej) 33. Z. Jia, et al.: Unity: Accelerating DNN Training Through Joint Optimization of Algebraic Transformations and Parallelization, OSDI, 2022.

gm799 (Gabriel) 34. J. Xing, et al.: Bolt: Bridging the Gap between Auto-tuners and Hardware-native Performance, MLSYS, 2022.

2024/11/27 Session 7: Optimisation in ML Compiler (max 14 mins presentation)

oea27 (Timi) 11. Z. Jia, O. Padon, J. Thomas, T. Warszawski, M. Zaharia, A. Aiken: TASO: Optimizing Deep Learning Computation with Automated Generation of Graph Substitutions: SOSP, 2019.

ak2599 (Andrew) 2. He, G., Parker, S., Yoneki, E.: X-RLflow: Graph Reinforcement Learning for Neural Network Subgraphs Transformation, MLSys, 2023.

wz337 (Andy) 12. KH. Wang, J. Zhai, M. Gao, Z. Ma, S. Tang, L. Zheng, Y. Li, K. Rong, Y. Chen, and Z. Jia: PET: Optimizing Tensor Programs with Partially Equivalent Transformations and Automated Corrections, ODSI, 2021.

hs899 (Hetong) 14. Y. Yang, et al.: Equality Saturation for Tensor Graph Superoptimization, MLSys, 2021.

sn666 (Sid) 4. Zheng, L. et al.: EinNet: Optimizing Tensor Programs with Derivation-Based Transformations, OSDI, 2023.

zc344 (Xavier) 18. L. Zheng, et al.: Ansor : Generating High-Performance Tensor Programs for Deep Learning, OSDI, 2020.
Paper 8 TVM: An Automated End-to-End Optimizing Compiler for Deep Learning and **Paper 9** TVM: End-to-End Compilation Stack for Deep Learning are also TVM papers, which are Ansor paper is an evolution of them. Thus, probably they need to be understood.

jkb55 (Kuba) 25. F. Kjolstad, et al.: The tensor algebra compiler, OOPSLA, 2017.