Big Data Systems – CS4545/CS6545
Winter 2021
Paper presentation schedule

How to read a paper - a good read

Notes

- Each group (2 persons) present 1 paper according to the schedule below.
  - For the presentation: prepare *PowerPoint slides with voice narration*, and upload to D2L by 12pm before the start of the class on the day of presentation.
  - This link provides instructions regarding how to create a slide-show with voice narration.
  - Each presentation should be around 18 minutes (not exceeding 20 minutes and not less than 15 minutes. Longer presentations will be stopped after 20 minutes.)
  - On the day of the presentation the presentation (*slides with voice narration)* will be played *online.* After a presentation, there will be an open discussion and students are encouraged to participate in it.
  - Attendance will be kept for the paper presentation days

- **Paper presentation schedule**

Date of presentation	**Category/ Topic**	**Paper**	**Presenters**
Feb 16	Storage	Lukas Vogel, Alexander van Renen, Satoshi Imamura, Viktor Leis, Thomas Neumann, Alfons Kemper. Mosaic: A Budget-Conscious Storage Engine for Relational Database Systems. VLDB 2020	David Thomson & Joshua Roberts
Feb 16	Storage	Jiajia Chu, Yunshan Tu, Yao Zhang, Chuliang Weng: Latte: A Native Table Engine On Nvme Storage. ICDE 2020	Juan Fernandez & Nithin Ivan
Feb 25	Indexing	Paolo Ferragina, Giorgio Vinciguerra. The PGM-index: a fully-dynamic compressed learned index with provable worst-case bounds. VLDB 2020	Eleanor McSporran & Guojun Tang
Feb 25	Indexing	Linwei Li, Kai Zhang, Jiading Guo, Wen He, Zhenying	Aditya Rambhatla

		He, Yinan Jing, Weili Han, X. Sean Wang: BinDex: A Two-Layered Index for Fast and Robust Scans. SIGMOD 2020	& Avinaba Mistry
Mar 11	Query processing	Michael Freitag, Maximilian Bandle, Tobias Schmidt, Alfons Kemper, Thomas Neumann. Adopting Worst-Case Optimal Joins in Relational Database Systems. VLDB 2020	Jeremy Robichaud & Mohammadali Rahnama
Mar 11	Indexing	Jialin Ding et al.: ALEX: An Updatable Adaptive Learned Index. SIGMOD 2020	Hadis Izadi Yekta
Mar 11	Query processing	Tim Gubner, Viktor Leis, Peter A. Boncz: Efficient Query Processing with Optimistically Compressed Hash Tables & Strings in the USSR. ICDE 2020	Ayoola Nurudeen Etiko & Iain Campbell
Mar 18	Parallel and distributed query processing	Rahul Potharaju et al. Helios: Hyperscale Indexing for the Cloud & Edge. VLDB 2020	Vishwa Barathy & Ryan Savoie
Mar 18	Parallel and distributed query processing	Jyoti Leeka, Kaushik Rajan. Incorporating Super-Operators in Big-Data Query Optimizers. VLDB 2020	Yogender Singh Dudee & Omkar Mangalgiri
Mar 18	Parallel and distributed query processing	Rundong Li, Wolfgang Gatterbauer, Mirek Riedewald: Near-Optimal Distributed Band-Joins through Recursive Partitioning. SIGMOD 2020	Geoffery Russo & Kevin Rail
Mar 25	Distributed and Cloud data management	Ahmed Alquraan, Alex Kogan, Virendra Marathe, Samer Al-Kiswany. Scalable, Near-Zero Loss Disaster Recovery for Distributed Data Stores. VLDB 2020	Nick Steeves & Jeremy Legere

Mar 25	Distributed and Cloud data management	Brad Glasbergen, Kyle Langendoen, Michael Abebe, Khuzaima Daudjee: ChronoCache: Predictive and Adaptive Mid-Tier Query Result Caching. SIGMOD 2020	Vishal Jain & Shermin Khosravi
Mar 25	Distributed and Cloud data management	Rebecca Taft et al.: CockroachDB: The Resilient Geo-Distributed SQL Database. SIGMOD 2020	Jharana Luitel & Shubham Verma
Mar 30	Beyond SQL	Christina Christodoulakis, Eric B Munson, Moshe Gabel, Angela Demke Brown, Renée J. Miller. Pytheas: Pattern-based Table Discovery in CSV Files. VLDB 2020	Nicholas Balcomb & Tristan Carrier
Mar 30	Beyond SQL	Devin Petersohn et al.: Towards Scalable Dataframe Systems. VLDB 2020	Marcus Kelly & Cormac Stewart
Mar 30	Supporting update intensive and mixed workloads	Christian Riegger, Tobias Vincon, Robert Gottstein and Ilia Petrov. MV-PBT: Multi-Version Indexing for Large Datasets and Mixed Workloads. EDBT 2020	Dylan Hubble & Timothy Meredith
Apr 8	Supporting update intensive and mixed workloads	Manos Athanassoulis, Kenneth Bøgh, Stratos Idreos. Optimal Column Layout for Hybrid Workloads. VLDB 2019	Anil Hitang & Tolulope Idris
Apr 8	Supporting update	Aritra Sengupta et al. Transactuations: Where Transactions Meet the Physical World. USENIX ATC 2019	Rojan Omidvar & David Keyes

	intensive and mixed workloads		
Apr 8	In-memory data management	Tiago R Kepe, Eduardo Cunha de Almeida, Marco A. Z. Alves. Database Processing-in-Memory: An Experimental Study. VLDB 2020	Jata MacCabe & Justin Beers
Apr 15	In-memory data management	Ajit Mathew and Changwoo Min. HydraList: A Scalable In-Memory Index Using Asynchronous Updates and Partial Replication. VLDB 2020	Bhanu Prakash Gude & Thomas Campbell
Apr 15	In-memory data management	Harald Lang, Alexander Beischl, Viktor Leis, Peter A. Boncz, Thomas Neumann, Alfons Kemper: Tree-Encoded Bitmaps. SIGMOD 2020	Bhargavi Oyonika & Samazder
Apr 15	Special topic: Blockchain data management	Hung Dang et al. Towards Scaling Blockchain Systems via Sharding. SIGMOD 2019	Naveen Kapoor & Jackson Dunn