List of Seminar Topics (Proposals)

Last update: Oct 20, 2023

Seminar Umbrella Topic

- **Papers in the fields of systems for data engineering, data management, and machine learning**

- **This Semester's Umbrella Topic: "Extensible Data Systems"**
  - Motivation
    - Meet requirements of emerging applications (multimedia, machine learning, domain-specific, …)
    - Enable the timely adoption of novel techniques and technologies (algorithms, hardware, …)
    - Handle the increasing specialization (data models, data representations, algorithms, hardware, …)
    - Facilitate research at systems and algorithms level
  - Active field of research for decades (at least 1980s till today)
  - Interesting challenges
    - Which abstractions are needed?
    - How to enable users to extend the system?
    - How to reach efficiency in spite of these abstractions?
  - Concepts have been proposed at all levels of the system stack

Seminar Topics (1/5)

- **Extensible Database Management Systems**
  - James Ong, Dennis Fogg, Michael Stonebraker:
    Implementation of Data Abstraction in the Relational Database System Ingres (SIGMOD Rec, 1984) [link]
  - Michael Stonebraker, Lawrence A. Rowe:
    The Design of POSTGRES (SIGMOD, 1986) [link]
  - Volker Linnemann, Klaus Küspert, Peter Dadam, Peter Pistor, R. Erbe, Alfons Kemper, Norbert Südkamp, Georg Walch, Mechtild Wallrath: **Design and Implementation of an Extensible Database Management System Supporting User Defined Data Types and Functions** (VLDB, 1988) [link]
  - Laura M. Haas, Walter Chang, Guy M. Lohman, John McPherson, Paul F. Wilms, George Lapis, Bruce G. Lindsay, Hamid Pirahesh, Michael J. Carey, Eugene J. Shekita:
    Starburst Mid-Flight: As the Dust Clears (IEEE Trans. Knowl. Data Eng., 1990) [link]
  - Peter A. Boncz, Martin L. Kersten: **MIL Primitives for Querying a Fragmented World** (VLDBJ, 1999) [link]
  - Markus Dreseler, Jan Kossmann, Martin Boissier, Stefan Klauck, Matthias Uflacker, Hasso Plattner:
    Hyrise Re-engineered: An Extensible Database System for Research in Relational In-Memory Data Management (EDBT, 2019) [link]

Seminar Topics (2/5)

- **Extensible Machine Learning Systems**
  - Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, Xiaoqiang Zheng: **TensorFlow: A System for Large-Scale Machine Learning** (OSDI, 2016) [link]
  - Shaoduo Gan, Xiangru Lian, Rui Wang, Jianbin Chang, Chengjun Liu, Hongmei Shi, Shengzhuo Zhang, Xianghong Li, Tengxu Sun, Jiawei Jiang, Binhang Yuan, Sen Yang, Ji Liu, Ce Zhang:
    BAGUA: Scaling up Distributed Learning with System Relaxations (PVLDB, 2022) [link]
- **Component-based Database/ML Systems**
  - Jacob D. Kahn, Vineel Pratap, Tatiana Likhomanenko, Qiantong Xu, Awni Y. Hannun, Jeff Cai, Paden Tomasello, Ann Lee, Edouard Grave, Gilad Avidov, Benoit Steiner, Vitaliy Liptchinsky, Gabriel Synnaeve, Ronan Collobert:
    Flashlight: Enabling Innovation in Tools for Machine Learning (PMLR, 2022) [link]
  - Immanuel Haffner, Jens Dittrich:
    mutable: A Modern DBMS for Research and Fast Prototyping (CIDR, 2023) [link]

Seminar Topics (3/5)

- **Optimizer Generators / Compiler Frameworks**
  - Don S. Batory, J. R. Barnett, Jorge F. Garza, K. P. Smith, K. Tsukuda, Brian C. Twichell, T. E. Wise: **GENESIS: An Extensible Database Management System** (IEEE Trans. Software Eng., 1988) [link]
  - Goetz Graefe, William J. McKenna: **The Volcano Optimizer Generator: Extensibility and Efficient Search** (ICDE, 1993) [link]
  - Chris Lattner, Mehdi Amini, Uday Bondhugula, Albert Cohen, Andy Davis, Jacques A. Pienaar, River Riddle, Tatiana Shpeisman, Nicolas Vasilache, Oleksandr Zinenko: **MLIR: Scaling Compiler Infrastructure for Domain Specific Computation** (CGO, 2021) [link]
- **Efficient Handling of User-defined Functions**
  - Fabian Hueske, Mathias Peters, Matthias Sax, Astrid Rheinländer, Rico Bergmann, Aljoscha Krettek, Kostas Tzoumas: **Opening the Black Boxes in Data Flow Optimization** (PVLDB, 2012) [link]
  - Florian Wolf, Iraklis Psaroudakis, Norman May, Anastasia Ailamaki, Kai-Uwe Sattler: **Extending database task schedulers for multi-threaded application code** (SSDBM, 2015) [link]
  - Viktor Rosenfeld, René Müller, Pinar Tözün, Fatma Özcan: **Processing Java UDFs in a C++ Environment** (SOCC, 2017) [link]

Seminar Topics (4/5)

- **Extensibility of Hardware Backends**
  - Holger Pirk, Oscar R. Moll, Matei Zaharia, Sam Madden: **Voodoo - A Vector Algebra for Portable Database Performance on Modern Hardware** (PVLDB, 2016) [link]
  - Tianqi Chen, Thierry Moreau, Ziheng Jiang, Lianmin Zheng, Eddie Q. Yan, Haichen Shen, Meghan Cowan, Leyuan Wang, Yuwei Hu, Luis Ceze, Carlos Guestrin, Arvind Krishnamurthy: **TVM: An Automated End-to-End Optimizing Compiler for Deep Learning** (OSDI, 2018) [link]
  - Annett Ungethüm, Johannes Pietrzyk, Patrick Damme, Alexander Krause, Dirk Habich, Wolfgang Lehner, Erich Focht: **Hardware-Oblivious SIMD Parallelism for In-Memory Column-Stores** (CIDR, 2020) [link]

Disclaimer: Co-authored by the lecturer

- **Extensible Data Analytics**
  - Xixuan Feng, Arun Kumar, Benjamin Recht, Christopher Ré: **Towards a unified architecture for in-RDBMS analytics** (SIGMOD, 2012) [link]
  - Tim Kraska, Ameet Talwalkar, John C. Duchi, Rean Griffith, Michael J. Franklin, Michael I. Jordan: **MLbase: A Distributed Machine-learning System** (CIDR, 2013) [link]

Seminar Topics (5/5)

- **Miscellaneous**
  - Patrick Damme, Annett Ungethüm, Johannes Pietrzyk, Alexander Krause, Dirk Habich, Wolfgang Lehner: **MorphStore: Analytical Query Engine with a Holistic Compression-Enabled Processing Model** (PVLDB, 2020) [link]
  - Moritz Sichert, Thomas Neumann:
    User-Defined Operators: Efficiently Integrating Custom Algorithms into Modern Databases (PVLDB, 2022) [link]
  - Xiaoying Wang, Weiyuan Wu, Jinze Wu, Yizhou Chen, Nick Zrymiak, Changbo Qu, Lampros Flokas, George Chow, Jiannan Wang, Tianzheng Wang, Eugene Wu, Qingqing Zhou:
    ConnectorX: Accelerating Data Loading From Databases to Dataframes (PVLDB, 2022) [link]
  - Michael Jungmair, André Kohn, Jana Giceva:
    Designing an Open Framework for Query Optimization and Compilation (PVLDB, 2022) [link] NEW
- **Alternative: Propose yet another topic/paper**
  - Should be related to the umbrella topic of extensible data systems

Disclaimer:
Co-authored by the lecturer