Project Ideas
CS 764, Fall 2021

This document contains a list of potential project topics. You are encouraged to explore your own ideas that interest you, which can be related to your own research.

1. Develop an educational database prototype for the undergraduate database class (CS 564) at UW Madison. Please contact the instructor if you are interested in this project.
- https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-830-database-systems-fall-2010/assignments/
- http://pages.cs.wisc.edu/~haiyun/cs564/buffermag_badgerDB/index.html

2. Modern OCC protocols demonstrate superior scalability and performance than 2PL concurrency control on multicore processors. Please identify the fundamental design principles that contribute to this effect and try to apply these design principles to 2PL for better scalability and performance. Please contact the instructor if you are interested in this project.

3. Conduct a survey of modern cloud-native serverless databases. What are the ideal features of a serverless database? Do existing serverless databases support all these features? If not, what features are missing and how can they be implemented? You may focus on either OLTP or OLAP but not necessarily both.

4. Modern cloud-native databases follow a storage-disaggregation architecture. How does this architecture fundamentally change the conventional wisdom of database design? You can explore the design space of an individual component of a database. Example components include consensus, concurrency control, logging, distributed commitment, data replication, query execution (e.g., join, group-by etc.), buffer management, etc.
- Yifei Yang, et al. *FlexPushdownDB: Hybrid Pushdown and Caching in a Cloud DBMS*, VLDB 2021
- Xiangyao Yu, et al. *PushdownDB: Accelerating a DBMS using S3 Computation*, ICDE 2020
- Junjay Tan, et al. *Choosing A Cloud DBMS: Architectures and Tradeoffs*, VLDB 2019

5. Cloud resources are highly elastic. With the same budget, a user may rent 1 server for 1 hour to finish a job, or 2 servers for 0.5 hour to finish the same job (assuming perfect scalability). This can speedup computation by $2\times$ for the same cost. Please investigate the cost aspect of cloud-native databases and identify optimization opportunities and challenges.
- Viktor Leis, Maximilian Kuschewski, *Towards Cost-Optimal Query Processing in the Cloud*, VLDB 2021