CS 5513 - Spring 2022 - Dr. Le Gruenwald
Group Projects (distributed 1/26/2022)

Project Categories:

1. Compare several existing techniques by means of simulation or prototyping and using benchmark/real data and queries/transactions for testing. You must justify your decision for choosing such techniques. **Your contribution: the comprehensive comparison study**.
2. Identify the deficiencies of existing techniques, derive your own technique and compare it with one or more existing ones by means of simulation or prototyping and using benchmark/real data and queries/transactions for testing. You must justify your decision for choosing those techniques for comparison. **Your contribution: your technique**.
3. Extend an existing open-source database system by prototyping additional components using existing techniques, incorporating the additional prototyped components into the open-source database system, and demonstrating the extended system's execution using data and queries/transactions from some applications. You must justify your decision for choosing such open-source system to extend and such components to prototype and how the extension will contribute to the database field. **Your contribution: the extended system with the additional components.**
4. Prototyping a system or some components of a system using existing techniques and demonstrating the system's execution using data and queries from some applications. You must justify your decision for choosing such a system/components of a system to prototype and how this prototype will contribute to the database field. **Your contribution: the prototyped system**.
5. Identify problems that no one has solved, derive a technique to solve them, and evaluate the performance of the technique using simulation or prototype. **Your contribution: being able to solve a problem that no one has solved before**.

Reading:

1. Daniel Adabi et al., "The Seattle Report on Database Research." ACM SIGMOD RECORD, Volume 48, Number 4, December 2019, pages 44-53, https://sigmodrecord.org/publications/sigmodRecord/1912/pdfs/07_Reports_Abadi.pdf
2. Daniel Abadi et al., "The Beckman Report on Database Research**",** ACM SIGMOD RECORD**,** Volume 43, Number 3, September 2014, pages 61-70, http://www.sigmod.org/publications/sigmod-record/1409/pdfs/09.reports-Abadi.pdf
3. Divyakant Agrawal et al., "Challenges and Opportunities with Big Data," Computing Research Association 2012, http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf.

Examples of Related Publication Sources: ACM International Conference on Management of Data (SIGMOD) and its associated workshops; International Conference on Very Large Data Bases (VLDB) and its associated workshops; IEEE International Conference on Data Engineering (ICDE) and its associated workshops; IEEE International Conference on Mobile Data Management (MDM) and its associated workshps; IEEE International Conference on Big Data (IEEE BigData) and its associated workshops; ACM Transactions on Database Systems (TODS); and IEEE Transactions on Knowledge and Data Engineering (TKDE).

Some possible project topics:

1. AI-Enabled Data Management:
   Using Machine Learning/Data Mining algorithms to perform database performance tuning autonomously (e.g. auto-indexing, auto-database partitioning), or to improve DBMS components (e.g. query optimization, failure recovery, database security)

2. Data Management for Cloud Computing, Green Computing, Big Data, and Modern Hardware:
    a. Green DBMS: energy-efficient and scalable DBMS components (query processing, concurrency control, replication, recovery, storage, indexing, etc.)
    b. DBMS components for cloud computing
    c. DBMS algorithms for big data
    d. Data security and privacy for cloud computing
    e. DBMS components for systems that use Flash memory
    f. Spatio-temporal query processing on high performance computing systems (e.g. clusters, GPGPU)
    g. Database partitioning algorithms for multi-tenant cloud databases
    h. Integration of big heterogeneous data

3. Stream and Sensor Network Databases
    a. Data management for fixed and mobile sensor network databases.
    b. Query processing in stream databases
    c. Privacy and security for stream databases
    d. Data integration in heterogeneous stream databases

4. Database and Crowdsourcing:
    a. Cleaning data from crowdsourcing
    b. Crowd database systems
    c. Declarative languages for crowdsourcing
    d. Utilization of social network information to support crowdsourcing.

5. Data Warehouse:
    a. Data cleaning/security/reliability
    b. Transaction processing
    c. Distributed data warehousing
    d. Design and maintenance of metadata repository
    e. View maintenance
    f. Indexing

6. Mobile Databases with/without energy consideration and with/without ad-hoc networks
    a. Transaction processing for homogeneous/heterogeneous systems
    b. Data Replication/Caching
    c. Data broadcast with/without security consideration
    d. Security and Privacy
    e. Mobile cloud data management.

7. Multimedia Database:
    a. Content-based search
    b. Indexing
    c. Data visualization

8. Other topics of your choice: static-distributed databases, real-time databases, temporal databases, web databases, geographic information systems, privacy-preserving data mining/data integration, etc.

SAMPLE CS 5513 PAST GROUP PROJECTS

1. Extending the open-source database system PostgreSQL to include a trajectory spatio-temporal query processing processor by implementing two different trajectory query processing algorithms and a trajectory query user-interface; and evaluating the extended system using Triage data and Microsoft mobile data. *(Project Category 3).*
2. Extending the open-source database system PostgreSQL to include additional indexing techniques and conducting experiments to compare all the indexing techniques in PostgreSQL (i.e. the indexing techniques existing in PostgreSQL as well as the additional indexing techniques). *(Project Category 3).*
3. Extending the existing data migration algorithm for MySQL to MongoDB and Analyzing its performance. *(Project Category 2).*
4. Enhancing the security of the REDIS database system by providing authentication and encryption. *(Project Category 3).*
5. Implementation and performance comparison of trajectory segmentation algorithms on GPU. *(Project Category 1).*
6. Extending Apache Cassandra: Attribute-Based Access Control. *(Project Category 3).*
7. Implementing an auto-indexing algorithm within an extension on PostgreSQL's platform. *(Project Category 3).*
8. Comparison of query performance in native DBMSs vs. SQL/MED: Wrapper interface access in PostgreSQL. *(Project Category 1).*
9. Developing and implementing a new algorithm to automate database schema design for NoSQL column store database systems, and evaluating its performance. *(Project Category 2).*
10. Developing a new transaction management technique for mobile multi-database computing environment, implementing the new algorithm and two existing algorithms, and conducting experiments to compare the new algorithm and the two existing algorithms. *(Project Category 2).*
11. Developing and prototyping a new algorithm to processing queries on data streams, prototyping two existing algorithms to process queries on data streams, and conducting experiments to compare the new algorithm with the two existing algorithms. *(Project Category 2).*
12. Comparing automatic database partitioning techniques for cloud databases by prototyping the three existing automatic database partitioning techniques in an Amazon cloud database and conducting experiments for comparison studies. *(Project Category 1).*