

CS 623 Spring 2023 Final Exam

RESEACH PRESENTATION IN NEW TRENDS IN DATABASES

Goal

The goal of this research project is to introduce you to a new and trendy topic in databases.

In this project, you will have to show that you understand the topic you chose to present.

You may need to show an illustrative example (code) covering the topic (necessary for certain topics).

You will produce a presentation and each group will present in class to explain your topic.

If you cannot present in class, make a 15-minute video of your presentation and submit.

You must make sure that you make the topic accessible to your classmates.

The videos will be available to the whole class and students will ask questions.

Organization

- To be done in a team of 2-3 students. You can use the same group as your projects.
- Each team will focus on different topics and conferences.
- Post your group and topic in the spreadsheet. Presentation is first come first serve.

Topics and conferences

Choose a topic in the list below:

- **Natural language interfaces:** Researchers are exploring the use of natural language interfaces to make databases more accessible to non-

technical users. The goal is to enable users to query databases using natural language instead of SQL.

- **Consistency Models in NoSQL and Cloud based databases:** Unlike traditional relational databases, many NoSQL databases sacrifice strong consistency in favor of high availability and partition tolerance. There is ongoing research into developing new consistency models that offer different levels of consistency guarantees.
- **Multiversion-Indexing:** It's a trending area of concurrent indexing. Traditional indexing techniques are designed to work with single-version databases, but with the rise of multi-version concurrency control (MVCC) systems, there is a need for new indexing techniques that can handle multiple versions of the same data. Researchers are exploring techniques such as Multi-Version B-Trees (MVB-Trees) and Optimistic Multi-Version Concurrency Control (OMVCC) to handle the challenges of indexing in MVCC systems.
- **Scalability of Massive amount of data and Apache Cassandra:** Research in this area aims to improve the scalability of Cassandra for handling increasingly large datasets. This includes exploring new partitioning strategies and load balancing techniques.
- **Blockchain:** Researchers are exploring the use of NoSQL databases and Couchbase in blockchain applications. This includes research into the development of new data models and algorithms that can be used to support decentralized and distributed applications.
- **Integration of Object-oriented databases with PostgreSQL:** Object-oriented databases (OODBs) and PostgreSQL are two different database management systems. While OODBs are designed to store and manage complex object-oriented data, PostgreSQL is a relational database management system that uses tables and relations to store data. However, there are ongoing research areas that explore the integration of OODBs with PostgreSQL such as in mapping objects to relational databases, object-relational hybrid databases.
- **Graph Query Languages:** There is ongoing research in developing efficient and expressive query languages for graph databases, including Neo4j. Some of the research is focused on extending existing query languages such as Cypher or GraphQL, while others are focused on developing new languages specifically for graph databases.

- **Machine Learning and Big data research in Distributed databases and HBase:** Distributed databases like HBase are designed to handle large-scale datasets, making them a natural fit for machine learning applications. Researchers are exploring new distributed machine learning algorithms that can scale to handle massive datasets in a distributed setting.
- **Performance Optimization of Hadoop and Map Reduce:** Hadoop and MapReduce have become popular tools for processing large-scale data, but there is ongoing research focused on improving their performance. This includes developing new algorithms for scheduling and task allocation, improving data locality, and optimizing network communication.

List of few conferences on databases:

- Massive Storage Systems and Technology Conference
 - <http://storageconference.us>
- ACM SIGMOD International Conference on Management of Data
 - <https://sigmod2019.org/>
- Very Large Databases Conference
 - <http://www.vldb.org/conference.html>
- Beyond Databases, Architectures and Structures (BDAS) Conference
 - <http://bdas.polsl.pl>
- Find others

Relevant Papers:

- 1.) "Mapping UML to Object-Relational Databases" by J. A. R. Nascimento and others. This paper proposes a new approach for mapping Unified Modeling Language (UML) models to object-relational databases, and evaluates its effectiveness using a case study.
- 2.) "Performance Optimization Techniques for Object-Relational Database Systems" by B. Wu and others. This paper presents a set of performance optimization techniques for object-relational database systems, including index selection, query optimization, and caching.
- 3.) Chrimes D, Zamani H. Using Distributed Data over HBase in Big Data Analytics Platform for Clinical Services. Comput Math Methods Med. 2017;2017:6120820. doi: 10.1155/2017/6120820. Epub 2017 Dec 11. PMID: 29375652; PMCID: PMC5742497.
- 4.) F. Abdelhedi, A. Ait Brahim, F. Atigui, and G. Zurfluh. 2017. MDA-based approach for NoSQL databases modelling. Lecture Notes in Computer Science (including

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

- 5.) de Sousa, V. M., & del Val Cura, L. M. (2018). Logical Design of Graph Databases from an Entity-Relationship Conceptual Model
- 6.) A. Chebotko, A. Kashlev and S. Lu, "A Big Data Modeling Methodology for Apache Cassandra," 2015 IEEE International Congress on Big Data, New York, NY, USA, 2015
- 7.) Vera-Olivera, H., Guo, R., Cruz Huacarpuma, R., et al. (2022). Data Modeling and NoSQL Databases - A Systematic Mapping Review. ACM Computing Surveys
- 8.) "Indexing Multiversion Databases" by Khaled Jouini and Geneviève Jomier, published in November 2007
- 9.) Bonifacio, A. S., Menolli, A. L. A., & Silva, F. (2014). Hadoop MapReduce Configuration Parameters and System Performance: a Systematic Review. In WorldComp2014-PDPTA2014 (pp. 1-6). Las Vegas, Nevada, USA.
- 10.) Gautam, C. S., Pandey, P., & Kumar, V. (2018). An improving query optimization process in Hadoop MapReduce using ACO-Genetic algorithm and HDFS map reduce Technique. International Journal of Computer Applications, 181(38)

Find other relevant papers or articles for the specific topic you choose.

References

Properly Cite your references. DO NOT CITE WIKEPEDIA!!!

General instructions

- The deliverables are a presentation, a review paper (6-10 pages) and a video(if you cannot present on 05/11)
- You will prepare 10-12 slides maximum on the chosen topic
 - Present the general research topic –
 - When did this research begin? Who are the leaders in this area of research? What are the leading conferences?
 - Why is it important? What are the applications? Etc.
 - What are the features of ...? Relevant Methodologies?
 - 1-3 slides will contain an illustrative example (e.g., example of graph database, postgresQL code etc.). All code must be tested. PROVIDE RELEVANT EXAMPLES ONLY.

- References (NO ONLINE LINKS, ONLY ACM, IEEE, AND JOURNAL PAPERS)
- Your presentation should be clear and concise. All code must be properly documented with.

Presentation

- Each team will have 10 minutes to present on Thursday (05/11). Teams that cannot present on that day should submit a video presentation 15 minutes long highlighting their work.
- Powerpoint slides, 10-15 slides long. 15 should be the max!

Paper and Video

- Each team will have 10 minutes to present on Thursday (05/11). Teams that cannot present on that day should submit a video presentation 15 minutes long highlighting their work.
- Write up a paper review of the research topic, 6-10 pages long, which explains a bit in-depth of the topic, its relevance, and your thoughts on the topic. Include any necessary methodologies/hypotheses and illustration examples or code.