Synergy between Quantum Computers and Databases

Valter Uotila

supervised by Prof. Jiaheng Lu and Prof. Jukka K. Nurminen
University of Helsinki, Finland

Abstract
Academia, industry, and societies are showing increasing interest in the possibilities of quantum computing. The research in the intersection of quantum computing and databases is still in its initial steps. This work represents several crucial data management and query processing problems that will benefit from quantum computing. We outline how quantum computing will tackle these challenges and what kind of outcomes and speed-ups we expect. We discuss the position of quantum computing in data management and raise awareness of possible security threats in encryption. We aim to be realistic and point out technical difficulties that currently restrict implementations.

1. Introduction

There are multiple different computing paradigms besides conventional CPU-based computing. Nowadays, the most exciting computing paradigm is quantum computing. It is based on quantum mechanics [1] although the modern quantum computing software [2, 3] can be used with almost no knowledge of quantum physics.

The quantum computers differ in their hardware. The most common hardware implementations are superconducting (IBM, Google, Rigetti), photonic (Xanadu), trapped ion (IonQ, Honeywell), adiabatic (D-wave), and silicon spin qubits (Intel, HRL). Amazon Braket, IBM Quantum, Xanadu, and D-wave Leap offer access to quantum computers and simulators in the cloud. The wide variety of hardware types shows that none of the types has yet become standard, and the competition between the quantum hardware companies is still going on. The future will show which quantum computing hardware type will become dominant.

Quantum computers will not take over classical computing. Instead, they will be computing units, like GPU processors or supercomputers, along with classical computers and databases. We can send them specific and computationally complex problems. Thus the hybrid approach will be the most realistic option for practical quantum computing.

1.1. Related work

There is relatively little research that applies quantum computing to databases and data management. Recent database research has applied quantum computing in transaction scheduling [4], and multiple query optimization [5]. The applicability of quantum computing on database query optimization is shortly outlined in [6].

Most of the discussion considers security issues [7, 8] which are also the most urgent questions. The National Institute of Standards and Technology (NIST) has started the post-quantum cryptography standardization process [9] because quantum computers will break RSA public-key cryptosystem in the future.

The future collapse of RSA and other classical encryption methods should also concern the database community. Databases and cloud infrastructures will need to implement quantum-safe encryption methods. Nevertheless, that does not mean that future security systems must be built on quantum computers. NIST evaluates encryption methods based on mathematical constructions considered too hard to be solved even on quantum computers.

2. Universal quantum circuit model and quantum annealing

We can roughly divide the quantum computing field into two parts depending on the hardware. These are NISQ (Noisy Intermediate-Scale Quantum) era hardware and quantum annealers. Algorithms for NISQ devices are implemented with the universal quantum circuit model.

The classical introduction to quantum circuit model-based quantum computing is [1]. Classical computing is based on bits, 0 and 1, whereas quantum computing is based on qubits

$$\varphi = \alpha|0\rangle + \beta|1\rangle,$$

where $\alpha, \beta \in \mathbb{C}$ are complex numbers whose norms satisfy $|\alpha|^2 + |\beta|^2 = 1$. The numbers α and β are called probability amplitudes. The requirement $|\alpha|^2 + |\beta|^2 = 1$ can be interpreted probabilistic way: the outcome $|0\rangle$ is determined by the amplitude α so that the probability of measuring $|0\rangle$ is $|\alpha|^2$. The probability of obtaining either $|0\rangle$ or $|1\rangle$ must be 1.

PhD Workshop of the 48th International Conference on Very Large Databases (VLDB 2022)
valter.uotila@helsinki (V. Uotila)

In general, a quantum system consists of multiple qubits. The system usually starts from the state $|00\cdots0\rangle$. The quantum computation proceeds by applying quantum logic gates to the system. Mathematically the gates are unitary matrices whose entries are complex numbers.

Finally, the system is measured and ends up in a state determined by the probabilities associated with the amplitudes. The measured outcome is the result of the algorithm, a sequence of classical bits.

Quantum annealing is similar to simulated annealing but performed on a quantum computer utilizing certain quantum mechanical phenomena. Quantum annealing is a heuristic method. To solve problems with quantum annealers, the problems are formulated in either the quadratic unconstrained binary optimization (QUBO) model or the Ising model, which are equivalent. As the formulations in [10] indicate, we can express many fundamental and NP-hard computer science problems using Ising and QUBO models. That shows strong evidence that many NP-hard problems in the database field can be solved using quantum computers.

The comprehensive introduction on formulating problems using Ising and QUBO models is [10]. Let x_i for $i = 1, \dots, n$ be binary variables. Then the objective function is

$$f(x) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} a_{i,j} x_i x_j + \sum_{i=1}^{n} b_i x_i, \tag{1}$$

where $x_i \in \{0,1\}$ and $a_{i,j}, b_i \in \mathbb{R}$ for $1 \leq i, j \leq n$. where coefficients $c_{i,j} \in \mathbb{R}$. The main goal of quantum annealing is to find a point x^* such that $f(x^*)$ is a global minimum of the objective function. The leading company building quantum annealers is D-wave.

3. Research directions

Although various optimization methods have been central topics of database research since the beginning, there is still much room for improvement. This section outlines research directions on optimizing databases with quantum computing.

3.1. Join order optimization

We propose a research direction to study the join order selection, a classical database optimization problem. The problem is relatively well-studied [11]. The classical solution to the problem uses dynamic programming. The current research has tackled the problem with deep reinforcement learning [12].

These kinds of combinatorically hard problems are the best match for quantum computing. Studying the problem using quantum computing, we aim to speed up the computation and improve the quality of the solution by extending the search space.

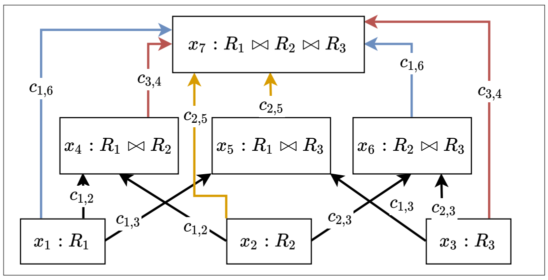

Figure 1: An example of the possibilities of joining three relations R_1, R_2 and R_3.

Figure 1 represents the join order problem for three relations R_1, R_2 and R_3. The problem fast becomes infeasible to solve exactly. Each edge carries a weight that describes the cost of joining the relations. The solution to the problem is a join order tree with a minimum total weight over all the join order trees. The tree must contain the single relations and the relation $R_1 \bowtie R_2 \bowtie R_3$.

When we express the problem with the QUBO model, we define the binary variables as the graph's vertices in Figure 1. The so-called couplers i.e. the variable pairs $x_i x_j$ in objective function (1) describe the edges. The coefficients $c_{i,j}$ in objective function (1) are determined by the cost function. The cost function estimates the cost of joining the relations. The solution's quality also depends on how accurately the cost function can estimate the joins. We impose constraints so that the result of the minimization problem is a valid join order tree whose total weight over the selected edges is the minimum. In the final model, we need auxiliary variables to produce correct results. We are also planning to utilize a hybrid approach since the number of variables and connectivity is large in the problem.

3.2. Size bound of conjunctive queries

Determining the size bound for conjunctive relational queries is another classical optimization problem for relational databases. The problem is defined and mostly solved in [13, 14]. The most complicated and general case with arbitrary functional dependencies stays open since it is strongly related to an unsolved and hard problem related to information-theoretical entropy vectors.

More precisely the problem is to determine an exponent $s(Q)$ for a conjunctive query $Q = R_0(u_0) \leftarrow R_1(u_1) \wedge \cdots \wedge R_n(u_m)$ in a database D which is given as a sequence of relations $R_1, \dots, R_n$. The size bound for the query Q in the database D is then

$$|Q(D)| \leq \mathrm{rmax}(D)^{s(Q)},$$

where $\mathrm{rmax}(D)$ is the largest relation in the database

D. The value $s(Q)$ is a solution to a linear program that maximizes the entropy of the variable u_0 concerning certain information theoretic constraints. The constraints arise from the functional dependencies in the database D. Unfortunately, to obtain the optimal solution, the linear program has infinitely many constraints and becomes infeasible. Since the solutions to the size bounds are represented as linear programs, there is already a known quantum algorithm [15] to solve them.

We believe that quantum computing will provide new aspects to modeling the problem. Because the problem is probabilistic and classically very hard, the quantum computing approach might produce good heuristic results and help understand the problem from a different angle. It is interesting to transfer the current solutions into a quantum computing format and research possible speed-ups and quality improvements.

3.3. Category theoretical methods for quantum computing and databases

In our previous research [16, 17] we have developed a conceptual framework to unify and model multi-model databases using category theory [18]. Our work has been a continuation of the research, which proves that relational databases are naturally category theoretical [19]. Category theory is a meta-mathematical field of mathematics, but it also has a lot of concrete applications, which are represented, for example, at the annual Applied Category Theory conference. Multi-model databases [20] are a particular class of databases whose single, integrated backend supports multiple data models and formats. In modern data management systems, it is crucial that we can handle a wide variety of data models seamlessly.

In [21] authors develop a comprehensive diagrammatic theory to model quantum processes and quantum computing. The approach is based on category theory. Since our previous research on multi-model databases [16, 17] is also based on category theory, this approach appears promising to connect the fields theoretically. Category theory as a high-level theoretical framework is needed since quantum computers are growing, and the gate-level design and coding of the algorithms are becoming harder. We need rigorous theoretical tools to model and design higher-level quantum computing systems, and the category theory-based approach seems to tackle the challenge well.

In [21, 22] authors use the category theory-based ideas to develop quantum natural language processing (QNLP) based on Lambek's pregroup grammars [23] and parametrized quantum circuits. We are working on applying these diagrammatic and category theory-based methods to represent SQL queries as parameterized quantum circuits. SQL is based on context-free grammar, equivalent to pregroup grammar in a certain sense. The parameterized circuits for SQL queries will be trained to estimate the query execution time, result's size, and cost. We are studying if the circuits will perform faster than the current estimation methods [24] and if they will provide better-quality estimations.

Natural continuation for this work is also express graph and document query languages with their diagrammatic representations and study query transformations between them. We have worked on data transformations [17] between relational, graph, and tree data models. Query transformations are the next step, and they are an essential part of multi-model databases. As category theory creates a consistent connection between schema and data [19], it will also create a solid connection between queries and instances. Because the query transformation depends on the database schema and instance, category theory will be a suitable theoretical framework for defining the transformations.

4. Conclusion

Academia, industry, and societies show growing interest in the possibilities of quantum computing. The optimization of data management and query processing will benefit from quantum computing frameworks. Although quantum computers are at an early stage, developing algorithms to tackle the presented challenges is essential. The database community should be aware of the security threats to the encryption that quantum computing eventually will create.

This work proposes multiple research directions to study join order optimization, determine the size bound of conjunctive queries, transform queries into parametrized circuits and perform query transformations using diagrammatic formalism. Quantum computers will solve countless problems in database research, and we are excited to be part of this new research area.

References

[1] M. A. Nielsen, I. L. Chuang, Quantum Computation and Quantum Information: 10th Anniversary Edition, Cambridge University Press, 2010. doi:10.1017/CBO9780511976667.

[2] M. Treinish, J. Gambetta, P. N. et al., Qiskit/qiskit: Qiskit 0.34.2, 2022. URL: https://doi.org/10.5281/zenodo.6027041. doi:10.5281/zenodo.6027041.

[3] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, K. McKiernan, J. J. Meyer, Z. Niu, A. Száva, N. Killoran, Pennylane: Automatic differentiation of hybrid quantum-classical computations, 2020. arXiv:1811.04968.

[4] T. Bittner, S. Groppe, Avoiding blocking by scheduling transactions using quantum annealing, in: Proceedings of the 24th Symposium on International Database Engineering & Applications, IDEAS '20, Association for Computing Machinery, New York, NY, USA, 2020. URL: https://doi.org/10.1145/3410566.3410593. doi:10.1145/3410566.3410593.

[5] I. Trummer, C. Koch, Multiple query optimization on the d-wave 2x adiabatic quantum computer, Proc. VLDB Endow. 9 (2016) 648–659. URL: https://doi.org/10.14778/2947618.2947621. doi:10.14778/2947618.2947621.

[6] M. Schönberger, Applicability of quantum computing on database query optimization, in: ACM SIGMOD Student Research Competition 2022 (undergraduate), 2022.

[7] W. Liu, P. Gao, Z. Liu, H. Chen, M. Zhang, A quantum-based database query scheme for privacy preservation in cloud environment, Security and Communication Networks 2019 (2019) 1–14. URL: http://dx.doi.org/10.1155/2019/4923590. doi:10.1155/2019/4923590.

[8] V. Giovannetti, S. Lloyd, L. Maccone, Quantum private queries, Phys. Rev. Lett. 100 (2008) 230502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.230502. doi:10.1103/PhysRevLett.100.230502.

[9] G. Alagic, J. Alperin-Sheriff, D. Apon, D. Cooper, Q. Dang, J. Kelsey, Y.-K. Liu, C. Miller, D. Moody, R. Peralta, R. Perlner, A. Robinson, D. Smith-Tone, Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process, NIST Internal or Interagency Report (NISTIR) 8309, 2020. URL: https://csrc.nist.gov/publications/detail/nistir/8309/final. doi:10.6028/NIST.IR.8309.

[10] A. Lucas, Ising formulations of many np problems, Frontiers in Physics 2 (2014). URL: http://dx.doi.org/10.3389/fphy.2014.00005. doi:10.3389/fphy.2014.00005.

[11] P. G. Selinger, M. M. Astrahan, D. D. Chamberlin, R. A. Lorie, T. G. Price, Access path selection in a relational database management system, in: Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data, SIGMOD '79, Association for Computing Machinery, New York, NY, USA, 1979, p. 23–34. URL: https://doi.org/10.1145/582095.582099. doi:10.1145/582095.582099.

[12] S. Krishnan, Z. Yang, K. Goldberg, J. Hellerstein, I. Stoica, Learning to optimize join queries with deep reinforcement learning, 2019. arXiv:1808.03196.

[13] A. Atserias, M. Grohe, D. Marx, Size bounds and query plans for relational joins, in: Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS '08, IEEE Computer Society, USA, 2008, p. 739–748. URL: https://doi.org/10.1109/FOCS.2008.43. doi:10.1109/FOCS.2008.43.

[14] G. Gottlob, S. T. Lee, G. Valiant, P. Valiant, Size and treewidth bounds for conjunctive queries, J. ACM 59 (2012). URL: https://doi.org/10.1145/2220357.2220363. doi:10.1145/2220357.2220363.

[15] A. W. Harrow, A. Hassidim, S. Lloyd, Quantum algorithm for linear systems of equations, Phys. Rev. Lett. 103 (2009) 150502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.150502. doi:10.1103/PhysRevLett.103.150502.

[16] V. Uotila, J. Lu, D. Gawlick, Z. H. Liu, S. Das, G. Pogossiants, Multicategory: Multi-model query processing meets category theory and functional programming, Proc. VLDB Endow. 14 (2021) 2663–2666. URL: https://doi.org/10.14778/3476311.3476314. doi:10.14778/3476311.3476314.

[17] V. Uotila, J. Lu, A formal category theoretical framework for multi-model data transformations, in: E. K. Rezig, V. Gadepally, T. Mattson, M. Stonebraker, T. Kraska, F. Wang, G. Luo, J. Kong, A. Dubovitskaya (Eds.), Heterogeneous Data Management, Polystores, and Analytics for Healthcare, Springer International Publishing, Cham, 2021, pp. 14–28.

[18] E. Riehl, Category Theory in Context, Aurora: Dover Modern Math Originals, Dover Publications, 2017. URL: https://math.jhu.edu/~eriehl/context.pdf.

[19] D. I. Spivak, Functorial data migration, 2013. arXiv:1009.1166.

[20] J. Lu, I. Holubová, Multi-model databases: A new journey to handle the variety of data, ACM Comput. Surv. 52 (2019). URL: https://doi.org/10.1145/3323214. doi:10.1145/3323214.

[21] B. Coecke, A. Kissinger, Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning, Cambridge University Press, 2017. doi:10.1017/9781316219317.

[22] R. Piedeleu, D. Kartsaklis, B. Coecke, M. Sadrzadeh, Open system categorical quantum semantics in natural language processing, arXiv:1502.00831 [cs, math] (2015). URL: http://arxiv.org/abs/1502.00831, arXiv: 1502.00831.

[23] J. Lambek, Type Grammar Revisited, volume 1582 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 1999, p. 1–27. URL: http://link.springer.com/10.1007/3-540-48975-4_1. doi:10.1007/3-540-48975-4_1.

[24] H. Lan, Z. Bao, Y. Peng, A survey on advancing the dbms query optimizer: Cardinality estimation, cost model, and plan enumeration, Data Science and Engineering 6 (2021) 86–101. doi:10.1007/s41019-020-00149-7.