Robust Query Processing: A Survey

Other titles in Foundations and Trends® in Databases

Extensible Query Optimizers in Practice
Bailu Ding, Vivek Narasayya and Surajit Chaudhuri
ISBN: 978-1-63828-452-9

Modern Techniques For Querying Graph-structured Databases
Amine Mhedhbi, Amol Deshpande and Semih Salihoğlu
ISBN: 978-1-63828-424-6

More Modern B-Tree Techniques
Goetz Graefe
ISBN: 978-1-63828-372-0

Consensus in Data Management: From Distributed Commit to Blockchain
Faisal Nawab and Mohammad Sadoghi
ISBN: 978-1-63828-160-3

Multidimensional Array Data Management
Florin Rusu
ISBN: 978-1-63828-148-1

Modern Datalog Engines
Bas Ketsman and Paraschos Koutris
ISBN: 978-1-63828-042-2

Robust Query Processing: A Survey

Jayant R. Haritsa
Indian Institute of Science
haritsa@iisc.ac.in

the essence of knowledge
Boston — Delft

Foundations and Trends® in Databases

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

J. R. Haritsa. *Robust Query Processing: A Survey*. Foundations and Trends® in Databases, vol. 15, no. 1, pp. 1–114, 2024.

ISBN: 978-1-63828-427-7

Editorial Scope

Foundations and Trends® in Databases publishes survey and tutorial articles in the following topics:

- Data Models and Query Languages
- Query Processing and Optimization
- Storage, Access Methods, and Indexing
- Transaction Management, Concurrency Control and Recovery
- Deductive Databases
- Parallel and Distributed Database Systems
- Database Design and Tuning
- Metadata Management
- Object Management
- Trigger Processing and Active Databases
- Data Mining and OLAP
- Approximate and Interactive Query Processing
- Data Warehousing
- Adaptive Query Processing
- Data Stream Management
- Search and Query Integration
- XML and Semi-Structured Data
- Web Services and Middleware
- Data Integration and Exchange
- Private and Secure Data Management
- Peer-to-Peer, Sensornet and Mobile Data Management
- Scientific and Spatial Data Management
- Data Brokering and Publish/Subscribe
- Data Cleaning and Information Extraction
- Probabilistic Data Management

Information for Librarians

Foundations and Trends® in Databases, 2024, Volume 15, 4 issues. ISSN paper version 1931-7883. ISSN online version 1931-7891. Also available as a combined paper and online subscription.

To my paternal and maternal grandfathers, Prof. T. S. Subbaraya and Prof. B. N. Balakrishna Rao, beacons of intellectual dedication.

To Seema, Tulasi and Sarangi, infusing life with melody.

Contents

Robust Query Processing: A Survey

Jayant R. Haritsa

Indian Institute of Science (IISc) Bangalore, India; haritsa@iisc.ac.in

---

ABSTRACT

The primordial function of a database system is to efficiently compute correct answers to user queries. Therefore, robust query processing (RQP), where strong numerical guarantees are provided on query performance, has been a long-standing core objective in the design of industrial-strength database engines. Unfortunately, however, RQP has proved to be a largely intractable and elusive challenge, despite sustained efforts spanning several decades. This problematic situation has arisen from a variety of knotty technical hurdles, including complex query representations, limited metadata coverage, coarse statistical models, and hypersensitive operator behaviors. Its impact is felt acutely since the performance degradation faced by database queries can be huge, reaching orders of magnitude as compared to an oracular ideal.

Notwithstanding this daunting history, the good news is that in recent times, there have been a host of exciting technical advances that collectively promise to materially address the robustness objective. The new approaches have been constructed at different levels in the database architecture, and tackle robustness in cost models, database operators, query execution plans and query processing strategies. Although most of this literature is based on statistical and geometric formulations, a significant corpus of machine learning-based techniques is also now available.

Jayant R. Haritsa (2024), "Robust Query Processing: A Survey", Foundations and Trends® in Databases: Vol. 15, No. 1, pp 1–114. DOI: 10.1561/1900000089.

In this monograph, we present an overview of these novel research paradigms, and highlight their strengths and limitations. Further, we enumerate a suite of open technical problems that remain to be solved to make RQP a contemporary reality.

1

Introduction

An organic reason for the ubiquitous popularity of database management systems is their support for *declarative* user queries, typically expressed in SQL. In this framework, the user only specifies the *end* objectives, leaving it to the database system to first identify and then execute the most efficient *means*, called "plan", to achieve these objectives. The identification and execution steps are performed by the *query optimizer* and *query executor* components, respectively, within the core of the database engine. Over the past half century, research on the design and implementation of these components has been a foundational topic for both the academic and industrial database communities.

A de facto global consensus exists on the technologies underlying most of the core database engine modules – for instance, two-phase locking (2PL) for concurrency control, write-ahead logging (WAL) for database recovery, least recently used (LRU-K) for memory management, and a combination of Bitmaps and B-trees for indexing database columns. Therefore, one might expect that a similar situation holds for query processing as well. However, despite the decades of research mentioned above, the unfortunate reality is that the proposed solutions have largely remained a "black art". This is due to the well-documented complexities and challenges of database query processing (Chaudhuri, 1998; Chaudhuri, 2009), which include complex query representations, limited metadata coverage, coarse statistical models, and hypersensitive operator behavior.

In fact, the prevalent situation is dire enough that a highly respected industry veteran was provoked to lament (Lohman, 2014): *The wonder isn't "Why did the optimizer pick a bad plan?" Rather, the wonder is "Why would the optimizer ever pick a decent plan?"*! He ended with the following exhortation to the research community: *"Let's attack problems that really matter, those that account for optimizer disasters, and stop polishing the round ball."* Similar sentiments have also been expressed by other academic and industrial database experts, including: *"Query optimizers do a terrible job of producing reliable, good plans (for complex queries) without a lot of hand tuning"* (Winslett, 2002), and *"Almost all of us who have worked on query optimization find the current state of the art unsatisfactory with known big gaps in the technology"* (Parameswaran, 2012).

What makes this parlous state of affairs particularly problematic is that the performance degradation faced by database queries can be *huge* – often in *orders of magnitude*, as compared to an oracular ideal that magically knows the correct inputs required for optimal query processing. As a case in point, when Query 19 of the TPC-DS benchmark (Transaction Processing Council, 2024a) is executed on a popular industrial-strength database system, the worst-case slowdown, relative to the hypothetical oracle, can exceed a *million*! (Dutt and Haritsa, 2016). Moreover, apart from the obvious detrimental impacts on user productivity and satisfaction, there are also adverse financial implications: the total cost of ownership is significantly increased due to over-provisioning, lost efficiency, and increased human administrative costs (Wiener *et al.*, 2009).

In the midst of this gloom and doom, the positive news is that in recent times there have been a host of exciting research advances, which collectively promise to provide strong foundations for designing the next generation of query processing engines. The new approaches have been constructed at different levels in the database architecture, and tackle robustness in cost models, database operators, query execution plans and query processing strategies. Although most of this literature is based on statistical and geometric formulations, a significant corpus of machine learning-based techniques has also now become visible.

The expectation is that these advances will eventually organically support **robust query processing (RQP)** with strong performance guarantees, relegating to the past the above-mentioned cynicism on this bedrock objective. Many of the new ideas owe their genesis to a series of influential and well-attended Dagstuhl Seminars on the topic of RQP over the past decade (Graefe *et al.*, 2010; Graefe *et al.*, 2012; Borovica-Gajic *et al.*, 2017; Böhm *et al.*, 2021). Further, they have arisen from research teams located at diverse locations across the world, including North America, Europe and Asia.

In this survey, we provide a holistic coverage of the RQP innovations, and highlight their strengths and limitations. Further, we enumerate a set of open technical problems and research directions that need to be studied to make RQP a contemporary reality.

1.1 Overview of Contents

The definition of robustness has itself been a subject of intense debate for a long time, and a consensus view has been difficult to achieve (Graefe *et al.*, 2010). For instance, if worst-case performance is improved at the expense of average-case performance, is that an acceptable notion of robustness? Or, would graceful degradation, as opposed to "performance cliffs", be the right perspective? Alternatively, is it the ability to seamlessly scale with workload complexity, database size and distributional skew? As yet another option, could we settle for providing strong theoretical guarantees relative to the oracular ideal? Perhaps, the real answer is that robustness encompasses all of these scenarios and more, with the specific choice being application-dependent.

The above semantic tangle is further complicated by the different levels at which notions of robustness can be introduced – for instance, at the granularity of individual *operators* (e.g. Borovica-Gajic *et al.*, 2018), or through entire query *plans* (e.g. Chu *et al.*, 1999), or over end-to-end query *executions* (e.g. Dutt and Haritsa, 2016). Moreover, one can take *algorithmic* (e.g. Tzoumas *et al.*, 2013), *statistical* (e.g. Wu *et al.*, 2013b) or *learning-based* (e.g. Malik *et al.*, 2007) approaches to incorporate the robustness features at individual levels.

In this monograph, we cover representative techniques along these various dimensions. Specifically, the survey is organized in the following sequence of sections:

Section 2: Background to Robust Query Processing We begin with an overview of declarative query optimization and processing. Then, we motivate the need for RQP and the systemic challenges faced in addressing this need. In particular, we focus on the two statistical models that fundamentally underlie the optimization process, namely, *operator cardinality estimation* and *operator cost estimation*. These models address orthogonal aspects of the data processing environment – the cardinality model captures the distributions and correlations present in the data, whereas the cost model reflects the behavior of the underlying hardware and physical operator implementations.

Section 3: Robust Operators Here, we consider robustness at the granularity of individual operators, with primary focus on the Scan and Join operations which carry out much of the heavy lifting in answering user queries. The basic idea is to design adaptive or unified operators that provide close to the best performance under all execution scenarios. Such an operator would make it unnecessary for the optimizer to choose between alternatives, thereby, by definition, eliminating erroneous choices.

Section 4: Robust Plans The next stage covers entire query plans, where we consider both strategies that are robust in expectation over a workload, and those that provide robustness on an individual query basis. The latter techniques leverage simple geometric assumptions on the behavior of *plan cost functions* – for instance, monotonicity with respect to predicate selectivity. We also look into how robustness can be effectively achieved by replacing the supposedly optimal plan with a mildly sub-optimal but stable alternative.

Section 5: Robust Query Execution We then move up from optimization to robust execution of entire queries. A key feature here is that the performance metrics are in comparison with the (offline) ideal – that is, the *lower bound*. This is a major conceptual shift from the norm in the earlier literature, where the comparison was always with the "upper bound", that is, the best among the alternative competing strategies available at the time.

Section 6: Structural Robustness Bounds The bounds in the previous section are dependent on the behavior of the database query optimizer over the parameter space. Here, we provide bounds that only depend on the *structure* of the parameter space, and not its *contents*. This quantum jump in robustness is achieved through the use of "spilling", wherein the outputs of intermediate operators in the plan tree are dropped on the ground, and not forwarded to the downstream nodes. The techniques presented here continue to leverage geometric assumptions on plan cost function behavior – specifically, in addition to monotonicity, both concavity and axis alignment are considered.

Section 7: Robust Cost Models While the operator cardinality model is the primary culprit for poor query performance, robustness can also be adversely impacted by errors in the operator cost model – we discuss mechanisms for addressing this problem in this section. In particular, we focus on how statistical approaches, augmented with careful calibration and focused sampling, can perform as well or even better than learning-based approaches, while being significantly more efficient wrt both training and inference.

Section 8: Machine Learning-based Techniques Learning-based approaches to RQP, which have been hotly pursued in recent years, are discussed here, covering both query-based and data-based techniques. The former is an example of supervised learning, with models constructed by training on a large set of queries and leveraging the actual cardinalities observed during execution as the labels. On the other hand, the data-based techniques fall under unsupervised learning, and model the joint probability density functions of the underlying data to capture distributions and correlations. Finally, there are hybrid models that leverage both queries and data in their learning process.

Section 9: Holistic Robustness Here, we show how the techniques discussed in the previous sections, which are at different layers in the database architecture, could be cohesively brought together in a complementary manner to maximize the overall system robustness.

Section 10: Future Research Directions Finally, we conclude with a suite of open research problems and future directions. The issues we pose include the impact of join-graph structure on robustness, creation of robustness benchmarks, and invoking ML techniques to determine when to use robust alternatives as opposed to their current avatars in database engines.

Overall, the big picture is that a rich variety of possibilities are currently available, and a judicious selection among them could lead to the desired robustness. Moreover, with the advent of the so-called Big Data world, wherein data is the engine driving virtually all aspects of human endeavor, the role of RQP assumes critical proportions.

1.2 Target Audience

Robust support for declarative query processing has been a long-standing concern for the database community, so we expect this monograph to be of broad relevance. In particular, the target audience for this monograph includes researchers, developers and students with an interest in the internals of database engines. The background expected is that of an introductory database systems course covering relational data models, declarative query languages, and basic query optimization and processing techniques.

Database researchers can expect the survey to provide fresh and radical perspectives on a classical research topic, serving to stimulate work on the further development of stable and efficient database engines. From the perspective of system developers and practitioners, the concepts and techniques presented in the monograph can serve as potent mechanisms for the redesign of their systems. Finally, for database instructors and students, the coverage will help in comprehending and appreciating the complexities and subtleties of industrial-strength query processing, going far beyond the toy examples typically covered in a classroom setting.

In particular, it may influence nascent PhD students looking for challenging research topics to cast their net beyond the middleware topics occupying center stage today – this is particularly important since the benefits of new engine technologies are automatically bestowed on all applications running on these platforms.

The primary source material for the monograph consists of the papers discussed in the various sections, complemented by supporting inputs from the rich corpus of literature on query optimization and processing. A sampling of relevant publications is given in the reference list, with emphasis on recent contributions to the field.

References

Aboulnaga, A. and S. Chaudhuri. (1999). "Self-tuning histograms: building histograms without looking at data". In: *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data. SIGMOD '99*. Philadelphia, Pennsylvania, USA: Association for Computing Machinery. 181–192. DOI: 10.1145/304182.304198.

Akdere, M., U. Çetintemel, M. Riondato, E. Upfal, and S. B. Zdonik. (2012). "Learning-based Query Performance Modeling and Prediction". In: *IEEE 28th International Conference on Data Engineering (ICDE 2012), Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012*. Ed. by A. Kementsietsidis and M. A. V. Salles. IEEE Computer Society. 390–401. DOI: 10.1109/ICDE.2012.64.

Avnur, R. and J. M. Hellerstein. (2000). "Eddies: Continuously Adaptive Query Processing". In: *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. SIGMOD '00*. Dallas, Texas, USA: Association for Computing Machinery. 261–272. DOI: 10.1145/342009.335420.

Babcock, B. and S. Chaudhuri. (2005). "Towards a Robust Query Optimizer: A Principled and Practical Approach". In: *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. SIGMOD '05*. Baltimore, Maryland: Association for Computing Machinery. 119–130. DOI: 10.1145/1066157.1066172.

Babu, S., P. Bizarro, and D. DeWitt. (2005). "Proactive Re-Optimization". In: *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. SIGMOD '05*. Baltimore, Maryland: Association for Computing Machinery. 107–118. DOI: 10.1145/1066157.1066171.

Bizarro, P., N. Bruno, and D. J. DeWitt. (2009). "Progressive Parametric Query Optimization". *IEEE Transactions on Knowledge and Data Engineering*. 21(04): 582–594. DOI: 10.1109/TKDE.2008.160.

Böhm, A., M. Christakis, E. Lo, and M. Rigger. (2021). "Ensuring the Reliability and Robustness of Database Management Systems (Dagstuhl Seminar 21442)". *Dagstuhl Reports*. 11(10): 20–35. DOI: 10.4230/DAGREP.11.10.20.

Borovica-Gajic, R., G. Graefe, and A. W. Lee. (2017). "Robust Performance in Database Query Processing (Dagstuhl Seminar 17222)". *Dagstuhl Reports*. 7(5): 169–180. DOI: 10.4230/DAGREP.7.5.169.

Borovica-Gajic, R., S. Idreos, A. Ailamaki, M. Zukowski, and C. Fraser. (2015). "Smooth Scan: Statistics-oblivious access paths". In: *31st IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea, April 13-17, 2015*. Ed. by J. Gehrke, W. Lehner, K. Shim, S. K. Cha, and G. M. Lohman. IEEE Computer Society. 315–326. DOI: 10.1109/ICDE.2015.7113294.

Borovica-Gajic, R., S. Idreos, A. Ailamaki, M. Zukowski, and C. Fraser. (2018). "Smooth Scan: robust access path selection without cardinality estimation". *VLDB J.* 27(4): 521–545. DOI: 10.1007/S00778-018-0507-8.

Chaudhuri, S. (1998). "An Overview of Query Optimization in Relational Systems". In: *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 1-3, 1998, Seattle, Washington, USA*. Ed. by A. O. Mendelzon and J. Paredaens. ACM Press. 34–43. DOI: 10.1145/275487.275492.

Chaudhuri, S. (2009). "Query optimizers: time to rethink the contract?" In: *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2009, Providence, Rhode Island, USA, June 29 - July 2, 2009*. Ed. by U. Çetintemel, S. B. Zdonik, D. Kossmann, and N. Tatbul. ACM. 961–968. DOI: 10.1145/1559845.1559955.

Chaudhuri, S., H. Lee, and V. R. Narasayya. (2010). "Variance Aware Optimization of Parameterized Queries". In: *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data. SIGMOD '10*. Indianapolis, Indiana, USA: Association for Computing Machinery. 531–542. DOI: 10.1145/1807167.1807226.

Chaudhuri, S., V. Narasayya, and R. Ramamurthy. (2008). "A Pay-as-You-Go Framework for Query Execution Feedback". *Proc. VLDB Endow.* 1(1): 1141–1152. DOI: 10.14778/1453856.1453977.

Chrobak, M., C. Kenyon, J. Noga, and N. E. Young. (2008). "Incremental Medians via Online Bidding". *Algorithmica*. 50(4): 455–478. DOI: 10.1007/S00453-007-9005-X.

Chu, F., J. Halpern, and J. Gehrke. (2002). "Least Expected Cost Query Optimization: What Can We Expect?" In: *Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. PODS '02*. New York, NY, USA: Association for Computing Machinery. 293–302. DOI: 10.1145/543613.543651.

Chu, F. C., J. Y. Halpern, and P. Seshadri. (1999). "Least Expected Cost Query Optimization: An Exercise in Utility". In: *Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 31 - June 2, 1999, Philadelphia, Pennsylvania, USA*. Ed. by V. Vianu and C. H. Papadimitriou. ACM Press. 138–147. DOI: 10.1145/303976.303990.

Cole, R. L. and G. Graefe. (1994). "Optimization of Dynamic Query Evaluation Plans". In: *Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data. SIGMOD '94*. Minneapolis, Minnesota, USA: Association for Computing Machinery. 150–160. DOI: 10.1145/191839.191872.

Deshpande, A., Z. Ives, and V. Raman. (2007). "Adaptive Query Processing". *Found. Trends databases*. 1(Jan.): 1–140. DOI: 10.1561/1900000001.

Doraiswamy, H., P. N. Darera, and J. R. Haritsa. (2007). "On the Production of Anorexic Plan Diagrams". In: *Proceedings of the 33rd International Conference on Very Large Data Bases, University of Vienna, Austria, September 23-27, 2007*. Ed. by C. Koch, J. Gehrke, M. N. Garofalakis, D. Srivastava, K. Aberer, A. Deshpande, D. Florescu, C. Y. Chan, V. Ganti, C. Kanne, W. Klas, and E. J. Neuhold. ACM. 1081–1092. URL: www.vldb.org/conf/2007/papers/research/p1081-d.pdf.

Doraiswamy, H., P. N. Darera, and J. R. Haritsa. (2008). "Identifying robust plans through plan diagram reduction". *Proc. VLDB Endow.* 1(1): 1124–1140. DOI: 10.14778/1453856.1453976.

Durkan, C. and C. Nash. (2019). "Autoregressive Energy Machines". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 1735–1744. URL: proceedings.mlr.press/v97/durkan19a.html.

Dutt, A. and J. R. Haritsa. (2016). "Plan Bouquets: A Fragrant Approach to Robust Query Processing". *ACM Trans. Database Syst.* 41(2): 11:1–11:37. DOI: 10.1145/2901738.

Dutt, A. and J. R. Haritsa. (2018). "Query Processing without Estimation". *Tech. rep.* No. TR-2014-01. DSL/SERC, Indian Institute of Science. URL: dsl.cds.iisc.ac.in/publications/report/TR/TR-2014-01.pdf.

Dutt, A., V. R. Narasayya, and S. Chaudhuri. (2017). "Leveraging Re-costing for Online Optimization of Parameterized Queries with Guarantees". In: *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD Conference 2017, Chicago, IL, USA, May 14-19, 2017*. Ed. by S. Salihoglu, W. Zhou, R. Chirkova, J. Yang, and D. Suciu. ACM. 1539–1554. DOI: 10.1145/3035918.3064040.

Dutt, A., C. Wang, V. R. Narasayya, and S. Chaudhuri. (2020). "Efficiently Approximating Selectivity Functions using Low Overhead Regression Models". *Proc. VLDB Endow.* 13(11): 2215–2228. URL: www.vldb.org/pvldb/vol13/p2215-dutt.pdf.

Dutt, A., C. Wang, A. Nazi, S. Kandula, V. R. Narasayya, and S. Chaudhuri. (2019). "Selectivity Estimation for Range Predicates using Lightweight Models". *Proc. VLDB Endow.* 12(9): 1044–1057. DOI: 10.14778/3329772.3329780.

Ganapathi, A., H. A. Kuno, U. Dayal, J. L. Wiener, A. Fox, M. I. Jordan, and D. A. Patterson. (2009). "Predicting Multiple Metrics for Queries: Better Decisions Enabled by Machine Learning". In: *Proceedings of the 25th International Conference on Data Engineering, ICDE 2009, March 29 2009 - April 2 2009, Shanghai, China.* Ed. by Y. E. Ioannidis, D. L. Lee, and R. T. Ng. IEEE Computer Society. 592–603. DOI: 10.1109/ICDE.2009.130.

Garey, M. R. and D. S. Johnson. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness (Series of Books in the Mathematical Sciences).* First Edition. W. H. Freeman. URL: www.amazon.com/Computers-Intractability-NP-Completeness-Mathematical-Sciences/dp/0716710455.

Germain, M., K. Gregor, I. Murray, and H. Larochelle. (2015). "MADE: Masked Autoencoder for Distribution Estimation". In: *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015.* Ed. by F. R. Bach and D. M. Blei. Vol. 37. *JMLR Workshop and Conference Proceedings.* JMLR.org. 881–889. URL: proceedings.mlr.press/v37/germain15.html.

Graefe, G. (2012). "New algorithms for join and grouping operations". *Comput. Sci. Res. Dev.* 27(1): 3–27. DOI: 10.1007/S00450-011-0186-9.

Graefe, G., W. Guy, H. A. Kuno, and G. N. Paulley. (2012). "Robust Query Processing (Dagstuhl Seminar 12321)". *Dagstuhl Reports.* 2(8): 1–15. DOI: 10.4230/DAGREP.2.8.1.

Graefe, G., A. C. König, H. A. Kuno, V. Markl, and K. Sattler. (2010). "Robust Query Processing (Dagstuhl Seminar 10381)". Dagstuhl Seminar Proceedings 10381. URL: drops.dagstuhl.de/portals/10381/.

Gray, J., A. Bosworth, A. Layman, and H. Pirahesh. (1996). "Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Total". In: *Proceedings of the Twelfth International Conference on Data Engineering, February 26 - March 1, 1996, New Orleans, Louisiana, USA*. Ed. by S. Y. W. Su. IEEE Computer Society. 152–159. DOI: 10.1109/ICDE.1996.492099.

Haas, P. J., J. F. Naughton, S. Seshadri, and A. N. Swami. (1996). "Selectivity and Cost Estimation for Joins Based on Random Sampling". *J. Comput. Syst. Sci.* 52(3): 550–569. DOI: 10.1006/JCSS.1996.0041.

Han, Y., Z. Wu, P. Wu, R. Zhu, J. Yang, L. W. Tan, K. Zeng, G. Cong, Y. Qin, A. Pfadler, Z. Qian, J. Zhou, J. Li, and B. Cui. (2021). "Cardinality Estimation in DBMS: A Comprehensive Benchmark Evaluation". *Proc. VLDB Endow.* 15(4): 752–765. DOI: 10.14778/3503585.3503586.

Harmouch, H. and F. Naumann. (2017). "Cardinality Estimation: An Experimental Survey". *Proc. VLDB Endow.* 11(4): 499–512. DOI: 10.1145/3186728.3164145.

Hasan, S., S. Thirumuruganathan, J. Augustine, N. Koudas, and G. Das. (2020). "Deep Learning Models for Selectivity Estimation of Multi-Attribute Queries". In: *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*. Ed. by D. Maier, R. Pottinger, A. Doan, W. Tan, A. Alawini, and H. Q. Ngo. ACM. 1035–1050. DOI: 10.1145/3318464.3389741.

Havenstein, D., P. Lysakovski, N. May, G. Moerkotte, and G. Steidl. (2020). "Fast Entropy Maximization for Selectivity Estimation of Conjunctive Predicates on CPUs and GPUs". In: *Proceedings of the 23rd International Conference on Extending Database Technology, EDBT 2020, Copenhagen, Denmark, March 30 - April 02, 2020*. Ed. by A. Bonifati, Y. Zhou, M. A. V. Salles, A. Böhm, D. Olteanu, G. H. L. Fletcher, A. Khan, and B. Yang. OpenProceedings.org. 546–554. DOI: 10.5441/002/EDBT.2020.65.

Hayek, R. and O. Shmueli. (2020a). "Improved Cardinality Estimation by Learning Queries Containment Rates". In: *Proceedings of the 23rd International Conference on Extending Database Technology, EDBT 2020, Copenhagen, Denmark, March 30 - April 02, 2020.* Ed. by A. Bonifati, Y. Zhou, M. A. V. Salles, A. Böhm, D. Olteanu, G. H. L. Fletcher, A. Khan, and B. Yang. OpenProceedings.org. 157–168. DOI: 10.5441/002/EDBT.2020.15.

Hayek, R. and O. Shmueli. (2020b). "NN-based Transformation of Any SQL Cardinality Estimator for Handling DISTINCT, AND, OR and NOT". *CoRR*. abs/2004.07009. arXiv: 2004.07009. URL: arxiv.org/abs/2004.07009.

Heimel, M., M. Kiefer, and V. Markl. (2015). "Self-Tuning, GPU-Accelerated Kernel Density Models for Multidimensional Selectivity Estimation". In: *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 - June 4, 2015.* Ed. by T. K. Sellis, S. B. Davidson, and Z. G. Ives. ACM. 1477–1492. DOI: 10.1145/2723372.2749438.

Hilprecht, B., A. Schmidt, M. Kulessa, A. Molina, K. Kersting, and C. Binnig. (2020). "DeepDB: Learn from Data, not from Queries!" *Proc. VLDB Endow.* 13(7): 992–1005. DOI: 10.14778/3384345.3384349.

Hulgeri, A. and S. Sudarshan. (2002). "Parametric Query Optimization for Linear and Piecewise Linear Cost Functions". In: *Proceedings of 28th International Conference on Very Large Data Bases, VLDB 2002, Hong Kong, August 20-23, 2002.* Morgan Kaufmann. 167–178. DOI: 10.1016/B978-155860869-6/50023-8.

Hüske, F. (2016). "Specification and optimization of analytical data flows". *PhD thesis*. Technical University of Berlin, Germany. URL: nbn-resolving.org/urn:nbn:de:101:1-201804165046.

Ioannidis, Y. E. and S. Christodoulakis. (1991). "On the Propagation of Errors in the Size of Join Results". In: *Proceedings of the 1991 ACM SIGMOD International Conference on Management of Data. SIGMOD '91.* Denver, Colorado, USA: Association for Computing Machinery. 268–277. DOI: 10.1145/115790.115835.

Jahangiri, S., M. J. Carey, and J.-C. Freytag. (2022). "Design trade-offs for a robust dynamic hybrid hash join". *Proc. VLDB Endow.* 15(10): 2257–2269. DOI: 10.14778/3547305.3547327.

Kabra, N. and D. J. DeWitt. (1998a). "Efficient Mid-Query Re-Optimization of Sub-Optimal Query Execution Plans". In: *SIGMOD 1998, Proceedings ACM SIGMOD International Conference on Management of Data, June 2-4, 1998, Seattle, Washington, USA*. Ed. by L. M. Haas and A. Tiwary. ACM Press. 106–117. DOI: 10.1145/276304.276315.

Kabra, N. and D. J. DeWitt. (1998b). "Efficient Mid-Query Re-Optimization of Sub-Optimal Query Execution Plans". In: *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data. SIGMOD '98*. Seattle, Washington, USA: Association for Computing Machinery. 106–117. DOI: 10.1145/276304.276315.

Karthik, S., J. R. Haritsa, S. Kenkre, and V. Pandit. (2016). "Platform-independent robust query processing". In: *32nd IEEE International Conference on Data Engineering, ICDE 2016, Helsinki, Finland, May 16-20, 2016*. IEEE Computer Society. 325–336. DOI: 10.1109/ICDE.2016.7498251.

Karthik, S., J. R. Haritsa, S. Kenkre, V. Pandit, and L. Krishnan. (2019). "Platform-Independent Robust Query Processing". *IEEE Trans. Knowl. Data Eng.* 31(1): 17–31. DOI: 10.1109/TKDE.2017.2664827.

Kiefer, M., M. Heimel, S. Breß, and V. Markl. (2017). "Estimating Join Selectivities using Bandwidth-Optimized Kernel Density Models". *Proc. VLDB Endow.* 10(13): 2085–2096. DOI: 10.14778/3151106.3151112.

Kipf, A., M. Freitag, D. Vorona, P. Boncz, T. Neumann, and A. Kemper. (2019a). "Estimating Filtered Group-By Queries is Hard: Deep Learning to the Rescue". In: *Proc of VLDB AIDB Workshop*. URL: drive.google.com/file/d/1Yt2w7_8UKFxsgcnWE8BjnFwhw509lyyY/view.

Kipf, A., T. Kipf, B. Radke, V. Leis, P. A. Boncz, and A. Kemper. (2019b). "Learned Cardinalities: Estimating Correlated Joins with Deep Learning". In: *9th Biennial Conference on Innovative Data Systems Research, CIDR 2019, Asilomar, CA, USA, January 13-16, 2019, Online Proceedings*. www.cidrdb.org. URL: cidrdb.org/cidr2019/papers/p101-kipf-cidr19.pdf.

Knuth, D. E. (1997). *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*. USA: Addison-Wesley Longman Publishing Co., Inc.

Kreyszig, E., H. Kreyszig, and E. J. Norminton. (2011). *Advanced Engineering Mathematics*. Tenth. Hoboken, NJ: Wiley.

Krishnan, S., Z. Yang, K. Goldberg, J. M. Hellerstein, and I. Stoica. (2018). "Learning to Optimize Join Queries With Deep Reinforcement Learning". *CoRR*. abs/1808.03196. arXiv: 1808.03196. URL: arxiv.org/abs/1808.03196.

Kumar, G. (2018). "Sub-query Based Approach for Robust Query Processing". *ME thesis*. CSA, IISc. URL: dsl.cds.iisc.ac.in/publications/thesis/gourav18.pdf.

Lehmann, C., P. Sulimov, and K. Stockinger. (2023). "Is Your Learned Query Optimizer Behaving As You Expect? A Machine Learning Perspective". DOI: 10.48550/ARXIV.2309.01551. arXiv: 2309.01551.

Leis, V., A. Gubichev, A. Mirchev, P. A. Boncz, A. Kemper, and T. Neumann. (2015). "How Good Are Query Optimizers, Really?" *Proc. VLDB Endow.* 9(3): 204–215. DOI: 10.14778/2850583.2850594.

Leis, V., B. Radke, A. Gubichev, A. Kemper, and T. Neumann. (2017). "Cardinality Estimation Done Right: Index-Based Join Sampling". In: *8th Biennial Conference on Innovative Data Systems Research, CIDR 2017, Chaminade, CA, USA, January 8-11, 2017, Online Proceedings*. www.cidrdb.org. URL: cidrdb.org/cidr2017/papers/p9-leis-cidr17.pdf.

Leis, V., B. Radke, A. Gubichev, A. Mirchev, P. A. Boncz, A. Kemper, and T. Neumann. (2018). "Query optimization through the looking glass, and what we found running the Join Order Benchmark". *VLDB J.* 27(5): 643–668. DOI: 10.1007/S00778-017-0480-7.

Li, G. (2010). "On the Design and Evaluation of a New Order-Based Join Algorithm". *MS thesis*. URL: asterix.ics.uci.edu/thesis/Guangqiang_Li_MS_thesis_2010.pdf.

Li, K. and G. Li. (2018). "Approximate Query Processing: What is New and Where to Go? - A Survey on Approximate Query Processing". *Data Sci. Eng.* 3(4): 379–397. DOI: 10.1007/S41019-018-0074-4.

Li, Z., O. Papaemmanouil, and M. Cherniack. (2016). "OptMark: A Toolkit for Benchmarking Query Optimizers". In: *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*. Ed. by S. Mukhopadhyay, C. Zhai, E. Bertino, F. Crestani, J. Mostafa, J. Tang, L. Si, X. Zhou, Y. Chang, Y. Li, and P. Sondhi. ACM. 2155–2160. DOI: 10.1145/2983323.2983658.

Liu, H., M. Xu, Z. Yu, V. Corvinelli, and C. Zuzarte. (2015). "Cardinality estimation using neural networks". In: *Proceedings of 25th Annual International Conference on Computer Science and Software Engineering, CASCON 2015, Markham, Ontario, Canada, 2-4 November, 2015*. Ed. by J. Gould, M. Litoiu, and H. Lutfiyya. IBM / ACM. 53–59. URL: dl.acm.org/citation.cfm?id=2886453.

Lohman, G. (2014). "Is Query Optimization a Solved Problem?" URL: wp.sigmod.org/?p=1075.

Lu, Y., S. Kandula, A. C. König, and S. Chaudhuri. (2021). "Pre-training Summarization Models of Structured Datasets for Cardinality Estimation". *Proc. VLDB Endow.* 15(3): 414–426. DOI: 10.14778/3494124.3494127.

Ma, L., B. Ding, S. Das, and A. Swaminathan. (2020). "Active Learning for ML Enhanced Database Systems". In: *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*. Ed. by D. Maier, R. Pottinger, A. Doan, W. Tan, A. Alawini, and H. Q. Ngo. ACM. 175–191. DOI: 10.1145/3318464.3389768.

Malik, T., R. C. Burns, and N. V. Chawla. (2007). "A Black-Box Approach to Query Cardinality Estimation". In: *Third Biennial Conference on Innovative Data Systems Research, CIDR 2007, Asilomar, CA, USA, January 7-10, 2007, Online Proceedings*. www.cidrdb.org. 56–67. URL: cidrdb.org/cidr2007/papers/cidr07p06.pdf.

Marcus, R., P. Negi, H. Mao, C. Zhang, M. Alizadeh, T. Kraska, O. Papaemmanouil, and N. Tatbul. (2019). "Neo: A Learned Query Optimizer". *Proc. VLDB Endow.* 12(11): 1705–1718. DOI: 10.14778/3342263.3342644.

Marcus, R. and O. Papaemmanouil. (2019). "Towards a Hands-Free Query Optimizer through Deep Learning". In: *9th Biennial Conference on Innovative Data Systems Research, CIDR 2019, Asilomar, CA, USA, January 13-16, 2019, Online Proceedings*. www.cidrdb.org. URL: cidrdb.org/cidr2019/papers/p96-marcus-cidr19.pdf.

Markl, V., V. Raman, D. Simmen, G. Lohman, H. Pirahesh, and M. Cilimdzic. (2004). "Robust Query Processing through Progressive Optimization". In: *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data. SIGMOD '04*. Paris, France: Association for Computing Machinery. 659–670. DOI: 10.1145/1007568.1007642.

Matias, Y., J. S. Vitter, and M. Wang. (1998). "Wavelet-Based Histograms for Selectivity Estimation". In: *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data. SIGMOD '98*. Seattle, Washington, USA: Association for Computing Machinery. 448–459. DOI: 10.1145/276304.276344.

Mattig, M., T. Fober, C. Beilschmidt, and B. Seeger. (2018). "Kernel-Based Cardinality Estimation on Metric Data". In: *Proceedings of the 21st International Conference on Extending Database Technology, EDBT 2018, Vienna, Austria, March 26-29, 2018*. Ed. by M. H. Böhlen, R. Pichler, N. May, E. Rahm, S. Wu, and K. Hose. OpenProceedings.org. 349–360. DOI: 10.5441/002/EDBT.2018.31.

Moerkotte, G., T. Neumann, and G. Steidl. (2009). "Preventing Bad Plans by Bounding the Impact of Cardinality Estimation Errors". *Proc. VLDB Endow.* 2(1): 982–993. DOI: 10.14778/1687627.1687738.

Müller, M., G. Moerkotte, and O. Kolb. (2018). "Improved Selectivity Estimation by Combining Knowledge from Sampling and Synopses". *Proc. VLDB Endow.* 11(9): 1016–1028. URL: www.vldb.org/pvldb/vol11/p1016-muller.pdf.

Muralikrishna, M. and D. J. DeWitt. (1988). "Equi-Depth Multidimensional Histograms". In: *Proceedings of the 1988 ACM SIGMOD International Conference on Management of Data. SIGMOD '88*. Chicago, Illinois, USA: Association for Computing Machinery. 28–36. DOI: 10.1145/50202.50205.

Negi, P., R. Marcus, H. Mao, N. Tatbul, T. Kraska, and M. Alizadeh. (2020). "Cost-Guided Cardinality Estimation: Focus Where it Matters". In: *36th IEEE International Conference on Data Engineering Workshops, ICDE Workshops 2020, Dallas, TX, USA, April 20-24, 2020*. IEEE. 154–157. DOI: 10.1109/ICDEW49219.2020.00034.

Neumann, T. and C. A. Galindo-Legaria. (2013). "Taking the Edge off Cardinality Estimation Errors using Incremental Execution". In: *Datenbanksysteme für Business, Technologie und Web (BTW), 15. Fachtagung des GI-Fachbereichs "Datenbanken und Informationssysteme" (DBIS), 11.-15.3.2013 in Magdeburg, Germany. Proceedings*. Ed. by V. Markl, G. Saake, K. Sattler, G. Hackenbroich, B. Mitschang, T. Härder, and V. Köppen. Vol. P-214. *LNI*. GI. 73–92. URL: https://dl.gi.de/handle/20.500.12116/17356.

Neumann, T. and B. Radke. (2018). "Adaptive Optimization of Very Large Join Queries". In: *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*. Ed. by G. Das, C. M. Jermaine, and P. A. Bernstein. ACM. 677–692. DOI: 10.1145/3183713.3183733.

Ngo, H. Q., C. Ré, and A. Rudra. (2014). "Skew strikes back: new developments in the theory of join algorithms". *SIGMOD Rec.* 42(4): 5–16. DOI: 10.1145/2590989.2590991.

Ngo, H. Q. (2018). "Worst-Case Optimal Join Algorithms: Techniques, Results, and Open Problems". In: *Proceedings of the 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems. PODS '18*. Houston, TX, USA: Association for Computing Machinery. 111–124. DOI: 10.1145/3196959.3196990.

Oord, A. van den, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu. (2016). "WaveNet: A Generative Model for Raw Audio". In: *The 9th ISCA Speech Synthesis Workshop, Sunnyvale, CA, USA, 13-15 September 2016*. ISCA. 125. URL: www.isca-speech.org/archive/SSW%5C_2016/abstracts/ssw9%5C_DS-4%5C_van%5C_den%5C_Oord.html.

Ortiz, J., M. Balazinska, J. Gehrke, and S. S. Keerthi. (2018). "Learning State Representations for Query Optimization with Deep Reinforcement Learning". In: *Proceedings of the Second Workshop on Data Management for End-To-End Machine Learning, DEEM@SIGMOD 2018, Houston, TX, USA, June 15, 2018*. Ed. by S. Schelter, S. Seufert, and A. Kumar. ACM. 4:1–4:4. DOI: 10.1145/3209889.3209890.

Ortiz, J., M. Balazinska, J. Gehrke, and S. S. Keerthi. (2019). "An Empirical Analysis of Deep Learning for Cardinality Estimation". *CoRR*. abs/1905.06425. arXiv: 1905.06425.

Ouared, A., A. Chadli, and M. A. Daoud. (2022). "DeepCM: Deep neural networks to improve accuracy prediction of database cost models". *Concurr. Comput. Pract. Exp.* 34(10). DOI: 10.1002/CPE.6724.

Parameswaran, A. G. (2012). "An interview with Surajit Chaudhuri". *XRDS*. 19(1): 38–39. DOI: 10.1145/2331042.2331055.

Park, Y., S. Zhong, and B. Mozafari. (2020). "QuickSel: Quick Selectivity Learning with Mixture Models". In: *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*. Ed. by D. Maier, R. Pottinger, A. Doan, W. Tan, A. Alawini, and H. Q. Ngo. ACM. 1017–1033. DOI: 10.1145/3318464.3389727.

Polyzotis, N. (2005). "Selectivity-Based Partitioning: A Divide-and-Union Paradigm for Effective Query Optimization". In: *Proceedings of the 14th ACM International Conference on Information and Knowledge Management. CIKM '05*. Bremen, Germany: Association for Computing Machinery. 720–727. DOI: 10.1145/1099554.1099730.

Purandare, S., S. Karthik, and J. R. Haritsa. (2018). "Dimensionality Reduction Techniques for Robust Query Processing". *Tech. rep.* No. TR-2018-02. DSL/CDS, Indian Institute of Science. URL: dsl.cds.iisc.ac.in/publications/report/TR/TR-2018-02.pdf.

Ramachandra, K., K. Park, K. V. Emani, A. Halverson, C. A. Galindo-Legaria, and C. Cunningham. (2017). "Froid: Optimization of Imperative Programs in a Relational Database". *Proc. VLDB Endow.* 11(4): 432–444. DOI: 10.1145/3186728.3164140.

Reddy, N. and J. R. Haritsa. (2005). “Analyzing Plan Diagrams of Database Query Optimizers”. In: *Proceedings of the 31st International Conference on Very Large Data Bases, Trondheim, Norway, August 30 - September 2, 2005*. Ed. by K. Böhm, C. S. Jensen, L. M. Haas, M. L. Kersten, P. Larson, and B. C. Ooi. ACM. 1228–1240. URL: www.vldb.org/archives/website/2005/program/paper/fri/p1228-reddy.pdf.

Rödiger, W., S. Idicula, A. Kemper, and T. Neumann. (2016). “Flow-Join: Adaptive skew handling for distributed joins over high-speed networks”. In: *32nd IEEE International Conference on Data Engineering, ICDE 2016, Helsinki, Finland, May 16-20, 2016*. IEEE Computer Society. 1194–1205. DOI: 10.1109/ICDE.2016.7498324.

Selinger, P. G., M. M. Astrahan, D. D. Chamberlin, R. A. Lorie, and T. G. Price. (1979). “Access Path Selection in a Relational Database Management System”. In: *Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data. SIGMOD ’79*. Boston, Massachusetts: Association for Computing Machinery. 23–34. DOI: 10.1145/582095.582099.

Silberschatz, A., H. F. Korth, and S. Sudarshan. (2020). *Database System Concepts, Seventh Edition*. McGraw-Hill Book Company. URL: https://www.db-book.com/.

State of New York. (2019). *Vehicle, snowmobile, and boat registrations.* URL: catalog.data.gov/dataset/vehicle-snowmobile-and-boat-registrations.

Stillger, M., G. M. Lohman, V. Markl, and M. Kandil. (2001). “LEO - DB2’s LEarning Optimizer”. In: *Proceedings of the 27th International Conference on Very Large Data Bases. VLDB ’01*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 19–28.

Sun, J. and G. Li. (2019). “An End-to-End Learning-based Cost Estimator”. *Proc. VLDB Endow.* 13(3): 307–319. DOI: 10.14778/3368289.3368296.

Sun, J., J. Zhang, Z. Sun, G. Li, and N. Tang. (2021). “Learned Cardinality Estimation: A Design Space Exploration and A Comparative Evaluation”. *Proc. VLDB Endow.* 15(1): 85–97. DOI: 10.14778/3485450.3485459.

Sun, L., C. Li, T. Ji, and H. Chen. (2023). "MOSE: A Monotonic Selectivity Estimator Using Learned CDF". *IEEE Trans. Knowl. Data Eng.* 35(3): 2823–2836. DOI: 10.1109/TKDE.2021.3112753.

Swami, A. (1989a). "Optimization of large join queries: combining heuristics and combinatorial techniques". In: *Proceedings of the 1989 ACM SIGMOD International Conference on Management of Data. SIGMOD '89*. Portland, Oregon, USA: Association for Computing Machinery. 367–376. DOI: 10.1145/67544.66961.

Swami, A. (1989b). "Optimization of large join queries: combining heuristics and combinatorial techniques". *SIGMOD Rec.* 18(2): 367–376. DOI: 10.1145/66926.66961.

Transaction Processing Council. (2024a). *TPC-DS*. URL: www.tpc.org/tpcds.

Transaction Processing Council. (2024b). *TPC-H*. URL: www.tpc.org/tpch.

Tzoumas, K., A. Deshpande, and C. S. Jensen. (2011). "Lightweight Graphical Models for Selectivity Estimation Without Independence Assumptions". *Proc. VLDB Endow.* 4(11): 852–863. URL: www.vldb.org/pvldb/vol4/p852-tzoumas.pdf.

Tzoumas, K., A. Deshpande, and C. S. Jensen. (2013). "Efficiently adapting graphical models for selectivity estimation". *VLDB J.* 22(1): 3–27. DOI: 10.1007/S00778-012-0293-7.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. Ed. by I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett. 5998–6008. URL: proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Venkatesh, S. K., J. R. Haritsa, S. Kenkre, and V. Pandit. (2018). "A Concave Path to Low-overhead Robust Query Processing". *Proc. VLDB Endow.* 11(13): 2183–2195. DOI: 10.14778/3275366.3275368.

Wang, X., C. Qu, W. Wu, J. Wang, and Q. Zhou. (2021a). "Are We Ready For Learned Cardinality Estimation?" *Proc. VLDB Endow.* 14(9): 1640–1654. DOI: 10.14778/3461535.3461552.

Wang, Y., C. Xiao, J. Qin, R. Mao, M. Onizuka, W. Wang, R. Zhang, and Y. Ishikawa. (2021b). "Consistent and Flexible Selectivity Estimation for High-Dimensional Data". In: *SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021*. Ed. by G. Li, Z. Li, S. Idreos, and D. Srivastava. ACM. 2319–2327. DOI: 10.1145/3448016.3452772.

Wang, Y. R., M. Willsey, and D. Suciu. (2023). "Free Join: Unifying Worst-Case Optimal and Traditional Joins". *Proc. ACM Manag. Data*. 1(2). DOI: 10.1145/3589295.

Wiener, J. L., H. A. Kuno, and G. Graefe. (2009). "Benchmarking Query Execution Robustness". In: *Performance Evaluation and Benchmarking, First TPC Technology Conference, TPCTC 2009, Lyon, France, August 24-28, 2009, Revised Selected Papers*. Ed. by R. O. Nambiar and M. Poess. Vol. 5895. *Lecture Notes in Computer Science*. Springer. 153–166. DOI: 10.1007/978-3-642-10424-4_12.

Winslett, M. (2002). "David DeWitt speaks out". *SIGMOD Rec.* 31(2): 50–62. DOI: 10.1145/565117.565127.

Wolf, F., M. Brendle, N. May, P. R. Willems, K. Sattler, and M. Grossniklaus. (2018a). "Robustness Metrics for Relational Query Execution Plans". *Proc. VLDB Endow.* 11(11): 1360–1372. DOI: 10.14778/3236187.3236191.

Wolf, F., N. May, P. R. Willems, and K. Sattler. (2018b). "On the Calculation of Optimality Ranges for Relational Query Execution Plans". In: *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*. Ed. by G. Das, C. M. Jermaine, and P. A. Bernstein. ACM. 663–675. DOI: 10.1145/3183713.3183742.

Woltmann, L., C. Hartmann, M. Thiele, D. Habich, and W. Lehner. (2019). "Cardinality estimation with local deep learning models". In: *Proceedings of the Second International Workshop on Exploiting Artificial Intelligence Techniques for Data Management, aiDM@SIGMOD 2019, Amsterdam, The Netherlands, July 5, 2019*. Ed. by R. Bordawekar and O. Shmueli. ACM. 5:1–5:8. DOI: 10.1145/3329859.3329875.

Wu, P. and G. Cong. (2021). "A Unified Deep Model of Learning from both Data and Queries for Cardinality Estimation". In: *SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021*. Ed. by G. Li, Z. Li, S. Idreos, and D. Srivastava. ACM. 2009–2022. DOI: 10.1145/3448016.3452830.

Wu, W. (2013). "Predicting Query Execution Time: Are Optimizer Cost Models Really Unusable?" URL: pages.cs.wisc.edu/~wentaowu/slides/ICDE-2013.pdf.

Wu, W., Y. Chi, H. Hacigümüs, and J. F. Naughton. (2013a). "Towards Predicting Query Execution Time for Concurrent and Dynamic Database Workloads". *Proc. VLDB Endow.* 6(10): 925–936. DOI: 10.14778/2536206.2536219.

Wu, W., Y. Chi, S. Zhu, J. Tatemura, H. Hacigümüs, and J. F. Naughton. (2013b). "Predicting query execution time: Are optimizer cost models really unusable?" In: *29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013*. Ed. by C. S. Jensen, C. M. Jermaine, and X. Zhou. IEEE Computer Society. 1081–1092. DOI: 10.1109/ICDE.2013.6544899.

Wu, W., J. F. Naughton, and H. Singh. (2016). "Sampling-Based Query Re-Optimization". In: *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*. Ed. by F. Özcan, G. Koutrika, and S. Madden. ACM. 1721–1736. DOI: 10.1145/2882903.2882914.

Wu, W., X. Wu, H. Hacigümüs, and J. F. Naughton. (2014). "Uncertainty Aware Query Execution Time Prediction". *Proc. VLDB Endow.* 7(14): 1857–1868. DOI: 10.14778/2733085.2733092.

Yang, Z., A. Kamsetty, S. Luan, E. Liang, Y. Duan, X. Chen, and I. Stoica. (2020). "NeuroCard: One Cardinality Estimator for All Tables". *Proc. VLDB Endow.* 14(1): 61–73. DOI: 10.14778/3421424.3421432.

Yang, Z., E. Liang, A. Kamsetty, C. Wu, Y. Duan, X. Chen, P. Abbeel, J. M. Hellerstein, S. Krishnan, and I. Stoica. (2019). "Deep Unsupervised Cardinality Estimation". *Proc. VLDB Endow.* 13(3): 279–292. DOI: 10.14778/3368289.3368294.

Yu, X., G. Li, C. Chai, and N. Tang. (2020). "Reinforcement Learning with Tree-LSTM for Join Order Selection". In: *36th IEEE International Conference on Data Engineering, ICDE 2020, Dallas, TX, USA, April 20-24, 2020*. IEEE. 1297–1308. DOI: 10.1109/ICDE48307.2020.00116.

Zaheer, M., S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola. (2017). "Deep Sets". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. Ed. by I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett. 3391–3401. URL: proceedings.neurips.cc/paper/2017/hash/f22e4747da1aa27e363d86d40ff442fe-Abstract.html.

Zhou, X., C. Chai, G. Li, and J. Sun. (2022). "Database Meets Artificial Intelligence: A Survey". *IEEE Trans. Knowl. Data Eng.* 34(3): 1096–1116. DOI: 10.1109/TKDE.2020.2994641.

Zhu, J., N. Potti, S. Saurabh, and J. M. Patel. (2017). "Looking ahead makes query plans robust: making the initial case with in-memory star schema data warehouse workloads". *Proc. VLDB Endow.* 10(8): 889–900. DOI: 10.14778/3090163.3090167.

Zhu, R., Z. Wu, Y. Han, K. Zeng, A. Pfadler, Z. Qian, J. Zhou, and B. Cui. (2021). "FLAT: Fast, Lightweight and Accurate Method for Cardinality Estimation". *Proc. VLDB Endow.* 14(9): 1489–1502. DOI: 10.14778/3461535.3461539.