Seminar: Query Optimization

Organizational Meeting

Chair of Database Systems
Chair of Data Science & Engineering
Garching, July 11, 2019

Overview

Weekly Meeting

- Mondays, 16:00 - 18:00, presumably starting Nov 07, 2019
- Room MI 02.09.014
- 2 presentations per meeting
- There will be an attendance log

Required Work

- Seminar paper (8̃ pages)
- Sample implementation (C++)
- Presentation (20 minutes + 10 minutes discussion)
- Moderate one discussion (act as the "devil's advocate", you should pair up for this)

Organization & Due Dates

Check in via email (radke@in.tum.de) or personally

1. Check in **soon after matching** for paper recommendations (preferences considered FCFS)
2. Check in when rough structure is planned
3. Check in when first draft is ready

Due Dates

- Structure: ca. 4 weeks prior to presentation date
- Presentation slides: 1 week prior to presentation date
- Seminar paper and sample implementation: 2 weeks after presentation date **(strict!)**

Topics - Join Ordering

- Greedy Operator Ordering [1, 2]
- IKKBZ [3, 4]
- Mixed Integer Linear Programming [5]
- Dynamic Programming [6, 7, 8]
- GroupBy Push Down [9, 10]
- Linearized DP [11, 12]
- Transformative Approaches [13]
- Iterative DP [14]
- Randomized Approaches [15, 16]

Topics - Cardinality Estimation

- Distinct Values [17, 18]
- Reservoir Sampling [19, 20, 21]
- Index-Based Join Sampling [22]
- Tighter Upper Bounds for Join Size Estimation [23]
- Learned Cardinalities [24, 25]

http://db.in.tum.de/teaching/ws1920/seminarAnfrageOpt

Bernhard Radke radke@in.tum.de

Have fun!

References I

[1] Leonidas Fegaras.

A new heuristic for optimizing large queries.

In Gerald Quirchmayr, Erich Schweighofer, and Trevor J. M. Bench-Capon, editors, *Database and Expert Systems Applications, 9th International Conference, DEXA '98, Vienna, Austria, August 24-28, 1998, Proceedings*, volume 1460 of *Lecture Notes in Computer Science*, pages 726–735. Springer, 1998.

[2] Nicolas Bruno, César A. Galindo-Legaria, and Milind Joshi.

Polynomial heuristics for query optimization.

In Feifei Li, Mirella M. Moro, Shahram Ghandeharizadeh, Jayant R. Haritsa, Gerhard Weikum, Michael J. Carey, Fabio Casati, Edward Y. Chang, Ioana Manolescu, Sharad Mehrotra, Umeshwar Dayal, and Vassilis J. Tsotras, editors, *Proceedings of the 26th International Conference on Data Engineering, ICDE 2010, March 1-6, 2010, Long Beach, California, USA*, pages 589–600. IEEE Computer Society, 2010.

[3] Toshihide Ibaraki and Tiko Kameda.

On the optimal nesting order for computing n-relational joins.

ACM Trans. Database Syst., 9(3):482–502, 1984.

References II

[4] Ravi Krishnamurthy, Haran Boral, and Carlo Zaniolo.
Optimization of nonrecursive queries.
In Wesley W. Chu, Georges Gardarin, Setsuo Ohsuga, and Yahiko Kambayashi, editors, *VLDB'86 Twelfth International Conference on Very Large Data Bases, August 25-28, 1986, Kyoto, Japan, Proceedings.*, pages 128–137. Morgan Kaufmann, 1986.

[5] Immanuel Trummer and Christoph Koch.
Solving the join ordering problem via mixed integer linear programming.
In Semih Salihoglu, Wenchao Zhou, Rada Chirkova, Jun Yang, and Dan Suciu, editors, *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD Conference 2017, Chicago, IL, USA, May 14-19, 2017*, pages 1025–1040. ACM, 2017.

[6] Patricia G. Selinger, Morton M. Astrahan, Donald D. Chamberlin, Raymond A. Lorie, and Thomas G. Price.
Access path selection in a relational database management system.
In Philip A. Bernstein, editor, *Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data, Boston, Massachusetts, USA, May 30 - June 1.*, pages 23–34. ACM, 1979.

References III

[7] Guido Moerkotte and Thomas Neumann.

Dynamic programming strikes back.

In Jason Tsong-Li Wang, editor, *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, pages 539–552. ACM, 2008.

[8] Guido Moerkotte, Pit Fender, and Marius Eich.

On the correct and complete enumeration of the core search space.

In Kenneth A. Ross, Divesh Srivastava, and Dimitris Papadias, editors, *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2013, New York, NY, USA, June 22-27, 2013*, pages 493–504. ACM, 2013.

[9] Marius Eich and Guido Moerkotte.

Dynamic programming: The next step.

In Johannes Gehrke, Wolfgang Lehner, Kyuseok Shim, Sang Kyun Cha, and Guy M. Lohman, editors, *31st IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea, April 13-17, 2015*, pages 903–914. IEEE Computer Society, 2015.

References IV

[10] Marius Eich, Pit Fender, and Guido Moerkotte.

Efficient generation of query plans containing group-by, join, and groupjoin.

VLDB J., 27(5):617–641, 2018.

[11] Thomas Neumann and Bernhard Radke.

Adaptive optimization of very large join queries.

In Gautam Das, Christopher M. Jermaine, and Philip A. Bernstein, editors, *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*, pages 677–692. ACM, 2018.

[12] Bernhard Radke and Thomas Neumann.

Lindp++: Generalizing linearized DP to crossproducts and non-inner joins.

In Torsten Grust, Felix Naumann, Alexander Böhm, Wolfgang Lehner, Theo Härder, Erhard Rahm, Andreas Heuer, Meike Klettke, and Holger Meyer, editors, *Datenbanksysteme für Business, Technologie und Web (BTW 2019), 18. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme"(DBIS), 4.-8. März 2019, Rostock, Germany, Proceedings*, volume P-289 of *LNI*, pages 57–76. Gesellschaft für Informatik, Bonn, 2019.

References V

[13] Arjan Pellenkoft, César A. Galindo-Legaria, and Martin L. Kersten.

The complexity of transformation-based join enumeration.

In Matthias Jarke, Michael J. Carey, Klaus R. Dittrich, Frederick H. Lochovsky, Pericles Loucopoulos, and Manfred A. Jeusfeld, editors, *VLDB'97, Proceedings of 23rd International Conference on Very Large Data Bases, August 25-29, 1997, Athens, Greece*, pages 306–315. Morgan Kaufmann, 1997.

[14] Donald Kossmann and Konrad Stocker.

Iterative dynamic programming: a new class of query optimization algorithms.

ACM Trans. Database Syst., 25(1):43–82, 2000.

[15] Florian Waas and Arjan Pellenkoft.

Join order selection - good enough is easy.

In Brian Lings and Keith G. Jeffery, editors, *Advances in Databases, 17th British National Conferenc on Databases, BNCOD 17, Exeter, UK, July 3-5, 2000, Proceedings*, volume 1832 of *Lecture Notes in Computer Science*, pages 51–67. Springer, 2000.

References VI

[16] César A. Galindo-Legaria, Arjan Pellenkoft, and Martin L. Kersten.

Uniformly-distributed random generation of join orders.

In Georg Gottlob and Moshe Y. Vardi, editors, *Database Theory - ICDT'95, 5th International Conference, Prague, Czech Republic, January 11-13, 1995, Proceedings*, volume 893 of *Lecture Notes in Computer Science*, pages 280–293. Springer, 1995.

[17] Moses Charikar, Surajit Chaudhuri, Rajeev Motwani, and Vivek R. Narasayya.

Towards estimation error guarantees for distinct values.

In Victor Vianu and Georg Gottlob, editors, *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, May 15-17, 2000, Dallas, Texas, USA*, pages 268–279. ACM, 2000.

[18] Michael J. Freitag and Thomas Neumann.

Every row counts: Combining sketches and sampling for accurate group-by result estimates.

In *CIDR 2019, 9th Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 13-16, 2019, Online Proceedings*. www.cidrdb.org, 2019.

References VII

[19] Jeffrey Scott Vitter.

Random sampling with a reservoir.

ACM Trans. Math. Softw., 11(1):37–57, 1985.

[20] Mohammed Al-Kateb, Byung Suk Lee, and Xiaoyang Sean Wang.

Adaptive-size reservoir sampling over data streams.

In *19th International Conference on Scientific and Statistical Database Management, SSDBM 2007, 9-11 July 2007, Banff, Canada, Proceedings*, page 22. IEEE Computer Society, 2007.

[21] Altan Birler.

Scalable reservoir sampling on many-core cpus.

In Peter A. Boncz, Stefan Manegold, Anastasia Ailamaki, Amol Deshpande, and Tim Kraska, editors, *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019.*, pages 1817–1819. ACM, 2019.

References VIII

[22] Viktor Leis, Bernhard Radke, Andrey Gubichev, Alfons Kemper, and Thomas Neumann.
Cardinality estimation done right: Index-based join sampling.
In *CIDR 2017, 8th Biennial Conference on Innovative Data Systems Research, Chaminade, CA, USA, January 8-11, 2017, Online Proceedings*. www.cidrdb.org, 2017.

[23] Walter Cai.
Tighter upper bounds for join cardinality estimates.
In Gautam Das, Christopher M. Jermaine, and Philip A. Bernstein, editors, *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*, pages 1805–1807. ACM, 2018.

[24] Andreas Kipf, Thomas Kipf, Bernhard Radke, Viktor Leis, Peter A. Boncz, and Alfons Kemper.
Learned cardinalities: Estimating correlated joins with deep learning.
In *CIDR 2019, 9th Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 13-16, 2019, Online Proceedings*. www.cidrdb.org, 2019.

References IX

[25] Lucas Woltmann, Claudio Hartmann, Maik Thiele, Dirk Habich, and Wolfgang Lehner.
Cardinality estimation with local deep learning models.
In Rajesh Bordawekar and Oded Shmueli, editors, *Proceedings of the Second International Workshop on Exploiting Artificial Intelligence Techniques for Data Management, aiDM@SIGMOD 2019, Amsterdam, The Netherlands, July 5, 2019*, pages 5:1–5:8. ACM, 2019.