

Jiaheng Lu
Professor, Supervisor for doctoral programme
Department of Computer Science
Unified DataBase Management System research group / Jiaheng Lu
Doctoral Programme in Computer Science
Postal address:
PL 68 (Pietari Kalmin katu 5)
00014
Finland
Email: jiaheng.lu@helsinki.fi
Phone: +358294151149



Qualifications

Principal Investigator

Period : 16.10.2015 - * in Department of Computer Science

Employment

Professor

Department of Computer Science
University of Helsinki
Finland
1 Oct 2020 → present

Unified DataBase Management System research group / Jiaheng Lu

University of Helsinki
1 Jan 2016 → present

Supervisor for doctoral programme

Doctoral Programme in Computer Science
University of Helsinki
Finland
1 Jan 2014 → present

Research outputs

A Survey on Mapping Semi-Structured Data and Graph Data to Relational Data

Yuan, G., Lu, J. & Yan, Z., Oct 2023, In: ACM Computing Surveys. 55, 10, 38 p.

Join Order Selection with Deep Reinforcement Learning: Fundamentals, Techniques, and Challenges

Yan, Z., Uotila, V. J. E. & Lu, J., Aug 2023, In: Proceedings of the VLDB Endowment. 16, 12, p. 3882-3885 4 p.

Quantum Computing for Databases: A Short Survey and Vision

Yuan, G., Lu, J., Chen, Y., Wu, S., Chang, Y., Yan, Z., Li, T. & Chen, G., Aug 2023, *Joint Workshops at 49th International Conference on Very Large Data Bases (VLDBW'23) — International Workshop on Quantum Data Science and Management (QDSM'23)*. CEUR, 8 p. (CEUR Workshop Proceedings; vol. 3462).

SimCost: cost-effective resource provision prediction and recommendation for spark workloads

Chen, Y., Hoque, M. A., Xu, P., Lu, J. & Tarkoma, S., 22 Jun 2023, In: Distributed and Parallel Databases. 30 p.

Quantum Machine Learning: Foundation, New Techniques, and Opportunities for Database Research

Winker, T., Groppe, S., Uotila, V. J. E., Yan, Z., Lu, J., Franz, M. & Mauerer, W., 5 Jun 2023, *Companion of the 2023 International Conference on Management of Data*. ACM, 8 p.

Multi-model query languages: taming the variety of big data

Guo, Q., Zhang, C., Zhang, S. & Lu, J., 31 May 2023, In: Distributed and Parallel Databases. 41 p.

Quantum Annealing Method for Dynamic Virtual Machine and Task Allocation in Cloud Infrastructures from Sustainability Perspective

Uotila, V. & Lu, J., 2023, *Proceedings - 2023 IEEE 39th International Conference on Data Engineering Workshops, ICDEW 2023*. Institute of Electrical and Electronics Engineers Inc., p. 105-110 6 p.

Effective Generation of Relational Schema from Multi-Model Data with Reinforcement Learning

Yuan, G., Lu, J. & Yan, Z., Oct 2022, *ER 2022: Conceptual Modeling*. Ralyte, J., Chakravarthy, S., Mohania, M., Jeusfeld, MA. & Karlapalem, K. (eds.). Springer, p. 224–235 12 p. (Lecture Notes in Computer Science (LNCS); vol. 13607).

Self-Adapting Design and Maintenance of Multi-Model Databases

Holubova, I., Koupil, P. & Lu, J., Sep 2022, *IDEAS '22: Proceedings of the 26th International Database Engineered Applications Symposium*. Desai, B. C. & Revesz, P. Z. (eds.). ACM, p. 9-15 7 p. (ACM International Conference Proceeding Series).

Automatic Performance Tuning for Distributed Data Stream Processing Systems

Herodotou, H., Odysseos, L., Chen, Y. & Lu, J., Aug 2022, *38TH IEEE International Conference on Data Engineering*. IEEE, (Data engineering).

Cross-Model Conjunctive Queries over Relation and Tree-structured Data

Chen, Y., Uotila, V., Lu, J., Liu, Z. H. & Das, S., 8 Apr 2022, *Database Systems for Advanced Applications. DASFAA 2022*. Bhattacharya, A., Li, JLM., Agrawal, D., Reddy, PK., Mohania, M., Mondal, A., Goyal & Kiran, RU. (eds.). Cham: Springer, p. 21-37 17 p. (Lecture Notes in Computer Science).

d-Simplexed: Adaptive Delaunay Triangulation or Performance Modeling and Prediction on Big Data Analytics

Chen, Y., Goetsch, P., Hoque, M. A., Lu, J. & Tarkoma, S., 1 Apr 2022, In: IEEE Transactions on Big Data. 8, 2, p. 458-469 12 p.

Storing Multi-model Data in RDBMSs based on Reinforcement Learning

Yuan, G., Lu, J., Zhang, S. & Yan, Z., Oct 2021, *CIKM '21: Proceedings of the 30th ACM International Conference on Information and Knowledge Management*. ACM, p. 3608–3611

A Formal Category Theoretical Framework for Multi-model Data Transformations

Uotila, V. J. E. & Lu, J., 20 Aug 2021, *Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB Workshops, Poly 2021 and DMAH 2021*. Rezig, E. K., Gadepally, V., Mattson, T., Stonebraker, M., Kraska, T., Wang, F., Luo, G., Kong, J. & Dubovitskaya, A. (eds.). Springer, Cham, p. 14-28

MultiCategory: multi-model query processing meets category theory and functional programming

Uotila, V. J. E., Lu, J., Gawlick, D., Liu, Z. H., Das, S. & Pogossiants, G., 20 Aug 2021, In: Proceedings of the VLDB Endowment. 14, 12, p. 2663–2666 4 p.

Automatic View Selection in Graph Databases

Zhang, C., Lu, J., Guo, Q., Zhang, X., Han, X. & Zhou, M., 6 Jul 2021, *SSDBM 2021: 33th International Conference on Scientific and Statistical Database Management*. ACM New York, p. 197–202 6 p.

Workload-Aware Performance Tuning for Autonomous DBMSs

Yan, Z., Lu, J., Chainani, N. & Lin, C., 19 Apr 2021, *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, p. 2365-2368 4 p. (IEEE International Conference on Data Engineering).

Holistic Evaluation in Multi-Model Databases Benchmarking

Zhang, C. & Lu, J., 2021, In: Distributed and Parallel Databases. 39, p. 1-33 33 p.

MORTAL: A Tool of Automatically Designing Relational Storage Schemas for Multi-model Data through Reinforcement Learning

Yuan, G. & Lu, J., 2021, *ER2021 Demos and Posters*. Lukyanenko, R., Samuel, B. & Sturm, A. (eds.). CEUR-WS.org, p. 7-12 (CEUR Workshop Proceedings; vol. 2958).

Performance Models of Data Parallel DAG Workflows for Large Scale Data Analytics

Shi, J. & Lu, J., 2021, *2021 IEEE 37TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING WORKSHOPS (ICDEW 2021)*. IEEE, p. 104-111 8 p. (IEEE International Conference on Data Engineering Workshops).

Quantum-Inspired Keyword Search on Multi-model Databases

Yuan, G., Lu, J. & SU, P., 2021, *International Conference on Database Systems for Advanced Applications: DASFAA 2021: Database Systems for Advanced Applications*. Cham: Springer, p. 585-602 (Lecture Notes in Computer Science ; vol. 12682).

Multi-Model Data Query Languages and Processing Paradigms

Guo, Q., Lu, J., Zhang, C., Sun, C. & Yuan, S., 20 Oct 2020, *CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. ACM New York, p. 3505-3506

Selectivity Estimation for Relation-Tree Joins

Zhang, C. & Lu, J., 19 Oct 2020, *SSDBM 2020: 32nd International Conference on Scientific and Statistical Database Management, Proceedings*. Pourabbas, E., Sacharidis, D., Stockinger, K. & Vergoulis, T. (eds.). ACM Digital library, p. 37-48 12 p.

One size does not fit all: accelerating OLAP workloads with GPUs

Zhang, Y., Zhang, Y., Lu, J., Wang, S., Liu, Z. & Han, R., 31 Jul 2020, In: *Distributed and Parallel Databases*. 38, p. 995-1037 43 p.

A Survey on Automatic Parameter Tuning for Big Data Processing Systems

Herodotou, H., Chen, Y. & Lu, J., Apr 2020, In: *ACM Computing Surveys*. 53, 2, p. 1-37 37 p., 43.

Multiple Set Matching with Bloom Matrix and Bloom Vector

Concas, F., Xu, P., Hoque, M. A., Lu, J. & Tarkoma, S., Mar 2020, In: *ACM Transactions on Knowledge Discovery from Data*. 14, 2, 21 p., 21.

Efficient string similarity join with taxonomy knowledge

Xu, P. & Lu, J., 1 Dec 2019, (Submitted) In: *Knowledge and Information Systems*.

Main-memory foreign key joins on advanced processors: design and re-evaluations for OLAP workloads

Zhang, Y., Zhang, Y., Zhou, X. & Lu, J., Dec 2019, In: *Distributed and Parallel Databases*. 37, 4, p. 469-506 38 p.

Cost-effective Resource Provisioning for Spark Workloads

Chen, Y., Lu, J., Chen, C., Hoque, M. A. & Tarkoma, S., 3 Nov 2019, *CIKM '19 : Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. New York, NY: ACM, p. 2477-2480 4 p.

Synergy of Database Techniques and Machine Learning Models for String Similarity Search and Join

Lu, J., Lin, C., Wang, J. & Li, C., Nov 2019, *CIKM '19 : Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. New York, NY: ACM, p. 2975-2976 2 p.

Fusion OLAP: Fusing the Pros of MOLAP and ROLAP Together for In-memory OLAP

Zhang, Y., Zhang, Y., Wang, S. & Lu, J., Sep 2019, In: *IEEE Transactions on Knowledge and Data Engineering*. 31, 9, p. 1722-1735 14 p.

Speedup Your Analytics: Automatic Parameter Tuning for Databases and Big Data Systems

Lu, J., Chen, Y., Herodotou, H. & Babu, S., 26 Aug 2019, In: Proceedings of the VLDB Endowment. 12, 12, p. 1970-1973 4 p.

PivotE: revealing and visualizing the underlying entity structures for exploration

Han, X., Chen, J., Lu, J., Chen, Y. & Du, X., Aug 2019, In: Proceedings of the VLDB Endowment. 12, 12, p. 1966-1969 4 p.

Towards a unified framework for string similarity joins

Xu, P. & Lu, J., 1 Jul 2019, In: Proceedings of the VLDB Endowment. 12, 11, p. 1289-1302 14 p.

Multi-model Databases: A New Journey to Handle the Variety of Data

Lu, J. & Holubová, I., Jul 2019, In: ACM Computing Surveys. 52, 3, 38 p., 55.

Crowd-Type: A Crowdsourcing-Based Tool for Type Completion in Knowledge Bases

Dong, Z., Tu, J., Fan, J., Lu, J., Du, X. & Ling, T. W., 2019, *Advances in Conceptual Modeling: ER 2018 Workshops Emp-ER, MoBiD, MREBA, QMMQ, SCME, Xi'an, China, October 22-25, 2018, Proceedings*. Woo, C., Lu, J., Li, Z., Ling, T. W., Li, G. & Lee, M. L. (eds.). Cham: Springer Nature Switzerland AG, p. 17-21 5 p. (Lecture Notes in Computer Science; vol. 11158).

Multi-model Database Management Systems - A Look Forward

Liu, Z. H., Lu, J., Gawlick, D., Helskyaho, H., Pogossiants, G. & Wu, Z., 2019, *Heterogeneous Data Management, Polystores, and Analytics for Healthcare : International Conference of Very Large databases VLDB 2018 Workshops, Poly and DMAH, Rio de Janeiro, Brazil, August 31, 2018, Revised Selected Papers*. Gadepally, V., Mattson, T., Stonebraker, M., Wang, F., Luo, G. & Teodoro, G. (eds.). Cham: Springer Nature Switzerland, p. 16-29 14 p. (Lecture Notes in Computer Science ; vol. 11470).

Neural Conversation Generation with Auxiliary Emotional Supervised Models

Zhou, G., Fang, Y., Peng, Y. & Lu, J., 2019, In: ACM Transactions on Asian Language Information Processing. 19, 2, 17 p., 19.

PivotE: Entity-Oriented Exploratory Search on Knowledge Graphs

Chen, J., Han, X., Chen, Y., Lu, J., Jacucci, G. & Ruotsalo, T., 2019, *Workshop on Data Systems for Interactive Analysis (DSIA)*. p. 1-6 6 p.

UDBMS: Road to Unification for Multi-model Data Management

Lu, J., Liu, Z. H., Xu, P. & Zhang, C., 2019, *Advances in Conceptual Modeling : ER 2018 Workshops Emp-ER, MoBiD, MREBA, QMMQ, SCME, Xi'an, China, October 22-25, 2018, Proceedings*. Woo, C., Lu, J., Li, Z., Ling, T. W., Li, G. & Lee, M. L. (eds.). Cham: Springer, p. 285-294 10 p. (Lecture Notes in Computer Science ; vol. 11158).

UniBench: A Benchmark for Multi-Model Database Management Systems

Zhang, C., Lu, J., Xu, P. & Chen, Y., 2019, *Performance Evaluation and Benchmarking for the Era of Artificial Intelligence: 10th TPC Technology Conference, TPCTC 2018, Rio de Janeiro, Brazil, August 27-31, 2018, Revised Selected Papers*. Nambiar, R. & Poess, M. (eds.). Cham: Springer, p. 7-23 16 p. (Lecture Notes in Computer Science; vol. 11135).

Unified management of multi-model data: (Vision paper)

Holubová, I., Svoboda, M. & Lu, J., 2019, *Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4-7, 2019, Proceedings*. Laender, A. H. F., Pernici, B., Lim, E-P. & Oliveira, J. P. M. D. (eds.). Cham: Springer Nature Switzerland, p. 439-447 9 p. (Lecture Notes in Computer Science ; vol. 11788).

Efficient Taxonomic Similarity Joins with Adaptive Overlap Constraint

Xu, P. & Lu, J., 15 Aug 2018, *CIKM '18 Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. New York, NY: ACM, p. 1563-1566 4 p.

Hierarchical Clustering of Complex Symbolic Data and Application for Emitter Identification

Xu, X., Lu, J. & Wang, W., Jul 2018, In: *Journal of Computer Science and Technology*. 33, 4, p. 807-822 16 p.

Optimal algorithms for selecting top-k combinations of attributes: theory and applications

Lin, C., Lu, J., Wei, Z., Wang, J. & Xiao, X., Feb 2018, In: *VLDB Journal*. 27, 1, p. 27-52 26 p.

Multi-model Databases and Tightly Integrated Polystores: Current Practices, Comparisons, and Open Challenges

Lu, J., Holubová, I. & Cautis, B., 2018, *CIKM '18 Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. New York, NY: ACM, p. 2301-2302 2 p.

Preface

Woo, C., Lu, J., Li, Z., Ling, T. W. & Li Z., L. T. W. (ed.), 2018, *Advances in Conceptual Modeling : ER 2018 Workshops Emp-ER, MoBiD, MREBA, QMMQ, SCME, Xi'an, China, October 22-25, 2018, Proceedings*. Woo, C., Lu, J., Li, Z., Ling, T. W., Li, G. & Lee, M. L. (eds.). Cham: Springer Nature Switzerland, p. V-VI 2 p. (Lecture Notes in Computer Science; vol. 11158).

Using Crowdsourcing for Fine-Grained Entity Type Completion in Knowledge Bases

Dong, Z., Fan, J., Lu, J., Du, X. & Ling, T. W., 2018, *Web and Big Data: Second International Joint Conference, APWeb-WAIM 2018, Macau, China, July 23-25, 2018, Proceedings, Part II*. Cai, Y., Ishikawa, Y. & Xu, J. (eds.). Cham: Springer International Publishing , p. 248-263 16 p. (Lecture Notes in Computer Science; vol. 10988).

Using hybrid algorithmic-crowdsourcing methods for academic knowledge acquisition

Dong, Z., Lu, J., Ling, T. W., Fan, J. & Chen, Y., Dec 2017, In: *Cluster Computing*. 20, 4, p. 3629-3641

Probesim: scalable single-source and top-k simrank computations on dynamic graphs

Liu, Y., Zheng, B., He, X., Wei, Z., Xiao, X., Zhen, K. & Lu, J., 1 Sep 2017, In: *Proceedings of the VLDB Endowment*. 11, 1, p. 14-26

Distributed Algorithms on Exact Personalized PageRank

Guo, T., Cao, X., Cong, G., Lu, J. & Lin, X., 9 May 2017, *SIGMOD '17: Proceedings of the 2017 ACM International Conference on Management of Data*. New York: ACM, p. 479-494 16 p.

Location-sensitive Query Auto-completion

Lin, C., Wang, J. & Lu, J., 3 Apr 2017, p. 819-820. 2 p.

Multi-model Data Management: What's New and What's Next?

Lu, J. & Holubová, I., 2017, *Proceeding of the 20th International Conference on extended Databases*. p. 602-605 4 p. (Proceeding of the 20th International Conference on extended Databases).

Top-k String Auto-Completion with Synonyms

Xu, P. & Lu, J., 2017, *Database Systems for Advanced Applications: 22nd International Conference, DASFAA 2017, Suzhou, China, March 27-30, 2017, Proceedings, Part II*. Candan, S., Chen, L., Pedersen, T. B., Chang, L. & Hua, W. (eds.). Cham: Springer International Publishing AG, Vol. 2. p. 202-218 (Lecture Notes in Computer Science; vol. 10178).

Towards Benchmarking Multi-Model Databases

Lu, J., 2017. 1 p.

Crowd-PANDA: Using Crowdsourcing Method for Academic Knowledge Acquisition

Dong, Z., Lu, J. & Ling, T. W., 2016, *Web-Age Information Management, Pt II*. Cui, B., Zhang, N., Xu, J., Lian & Liu, D. (eds.). Springer, p. 531-533 3 p. (Lecture Notes in Computer Science; vol. 9659).

Incremental Hierarchical Clustering of Stochastic Pattern-Based Symbolic Data

Xu, X., Lu, J. & Wang, W., 2016, *Advances in Knowledge Discovery and Data Mining : 20th Pacific-Asia Conference, PAKDD 2016, Auckland, New Zealand, April 19-22, 2016, Proceedings, Part II*. Khan, L., Bailey, J., Washio, T., Dobbie, G., Huang, J. Z. & Wang, R. (eds.). Cham: Springer International Publishing AG, p. 156-167 12 p. (Lecture Notes in Computer Science ; vol. 9652).

PANDA: A Platform for Academic Knowledge Discovery and Acquisition

Dong, Z., Lu, J. & Ling, T. W., 2016, *2016 International Conference on Big Data and Smart Computing, BigComp 2016, Hong Kong, China, January 18-20, 2016*. IEEE Computer Society, p. 10-17 8 p.

Boosting the Quality of Approximate String Matching by Synonyms

Lu, J., Lin, C., Wang, W., Li, C. & Xiao, X., 2015, In: *ACM Transactions on Database Systems*. 40, 3, 42 p., 15.

Towards Maximum Independent Sets on Massive Graphs

Liu, Y., Lu, J., Yang, H., Xiao, X. & Wei, Z., 2015, In: *Proceedings of the VLDB Endowment*. 8, 13, p. 2122-2133 12 p.

Activities

Multi-model Query Processing Meets Category Theory and Functional Programming

Valter Johan Edvard Uotila (Speaker), Jiaheng Lu (Speaker), Dieter Gawlick (Speaker), Zhen Hua Liu (Speaker), Souripriya Das (Speaker) & Gregory Pogossiants (Speaker)
20 Aug 2021

46th International Conference on Very Large Data Bases - VLDB2020 (Event)

Jiaheng Lu (Member)
Mar 2019 → Mar 2020

The 37th International Conference on Conceptual Modeling

Jiaheng Lu (Member of organizing committee)
22 Oct 2018 → 25 Oct 2018

35th IEEE International Conference on Data Engineering (ICDE 2019) (Event)

Jiaheng Lu (Board member)
1 Jul 2018 → 31 Dec 2018

Springer (Publisher)

Jiaheng Lu (Reviewer)
2018

APWeb-WAIM 2017 (Event)

Jiaheng Lu (Board member)
2017 → 2018

Australia Database Conference (Event)

Jiaheng Lu (Board member)
2017 → 2018

The fourth workshop on keyword search and data exploration

Jiaheng Lu (Chair)
20 May 2016

International Conference on Data Engineering (Event)

Jiaheng Lu (Board member)
2016 → 2017

International Conference on Very Large Data Bases (Event)

Jiaheng Lu (Board member)
2016 → 2017

ACM-Sigmod International Conference on Management of Data (Event)
Jiaheng Lu (Board member)
2015 → 2016