Feng "George" Yu

Associate Professor
Computer Science and Information Systems
Youngstown State University
Youngstown, OH, 44555

Email: fyu@ysu.edu
Office: +1 (330) 941-1775
Homepage: http://fyu.people.ysu.edu
Research Lab: http://datalab.ysu.edu/

Research Interests: Database Systems, Big Data, Data Analytics, Cloud Computing

1 Professional Experience

Youngstown State University, Youngstown, OH	Associate Professor	2018-current
Youngstown State University, Youngstown, OH	Assistant Professor	2013-2018

2 Education

Southern Illinois University, Carbondale, IL	Computer Science	Ph.D.	2013
Shandong University, China	Mathematics	M.S.	2008
Northeastern University, China	Information and Computation Science	B.S.	2005

3 Publication

3.1 Peer-Reviewed Journal Publications

17. **F. Yu**, S. Cal, E. Cheng, L. Kerns, W.-D. Xiong, Non-Parametric Error Estimation for σ-AQP using Optimized Bootstrap Sampling, *International Journal of Computers and Their Applications*, 29 (1), 38-47, 2022

16. **F. Yu**, D.S. Wilson, T.M. Wells, M.A. Hamdi, X. Huang, W.-C. Hou, Scalable Correlated Sampling for Join Query Estimations on Big Data, *International Journal of Computers and Their Applications*, 27 (1), 14-23, 2020.

15. S. Clark, **F. Yu**, L. Gu, X.-J. Min, Expanding Alternative Splicing Identification by Integrating Multiple Sources of Transcription Data in Tomato, *Frontiers in Plant Science*, Vo. 10, 12 pages, May 2019 (DOI: 10.3389/fpls.2019.00689)

14. R. M. Almutairi, M. Hamdi, **F. Yu**, W.-C. Hou, An Evaluation of the Performance of Join Core and Join Indices Query Processing Methods, *International Journal of Computers and Their Applications*, Vol. 25, No. 3, pages 123-131, September 2018.

13. **F. Yu**, T. J. Matacic, B. J. Latronica, OB-Tree: A New Write Optimization Index on Out-of-Core Column-Store Databases. *International Journal of Intelligent Information and Database Systems*, Vol. 11, No. 1, pages 46-66, 2018 (DOI: 10.1504/IJIIDS.2018.10012700)

12. G. Sablok, B. Powell, J. Braessler, **F. Yu**, X.-J. Min, Comparative landscape of alternative splicing in fruit plants, *Current Plant Biology*, July 2017 (DOI: 10.1016/j.cpb.2017.06.001)

11. M. Hamdi, W. Xiong, **F. Yu**, S. Alswedani, W.-C. Hou, An Aggressive Concurrency Control Protocol for Main Memory Databases, *International Journal of Computer Applications*, Vol. 155, No. 2, pages 7-13, December 2016

10. B. Powell, V. Amerishetty, J. Meinken, G. Knott, **F. Yu**, C. Cooper, X.-J. Min, ProtSecKB: The Protist Secretome and Subcellular Proteome, *Computational Molecular Biology*, Vol. 6, No. 4, pages 1-12, December 2016. (DOI:10.5376/cmb.2016.06.0004)

9. W. Xiong, M. Hamdi, **F. Yu**, W.-C. Hou, A Prudent-Precedence Concurrency Control Protocol for High Data Contention Database Environments, *International Journal of Database Management Systems (IJDMS)*, Vol. 8, No. 5, pages 1-14, October 2016

8. **F. Yu**, E. S. Jones, W. Xiong, M. Hamdi, W.-C. Hou, An Asynchronous Method for Write Optimization of Column-Store Databases in Map-Reduce, *International Journal of Big Data*, Vol. 2, No. 4, pages 1-9, 2015.

7. X. J. Min, B. Powell, J. Braessler, J. Meinken, **F. Yu**, and G. Sablok, Genome-Wide Cataloging and Analysis of Alternatively Spliced Genes in Cereal Crops, *BMC Genomics*, Vol. 16, No. 1, 13 pages, December 2015.

6. **F. Yu**, C. Stella, K. A. Schueller, A Design of Heterogeneous Cloud Infrastructure for Big Data and Cloud Computing Services, *Open Journal of Mobile Computing and Cloud Computing*, Vol 1, No. 2, pages 1-16, November, 2014

5. **F. Yu**, W.-C. Hou, C. Luo, Estimation Error Analysis of Range Selection Queries using Histogram and Sample in Low Dimensional Spaces, *International Journal of Computers and Their Applications*, Vol. 21, No. 3, pages 151-159, September, 2014

4. **F. Yu**, J. Zhang, W.-C. Hou, C. Luo, M. Wainer, A. Pabbaraju, Enhancing SQL with Set-Comparison Operators, *International Journal of Computer and Information Technology*, Vol. 3, Issue 3, pages 590-595, May 2014

3. **F. Yu**, W.-C. Hou, M. Wainer, D. Che, C. Luo, Sufficient Statistics for Re-optimizing Repetitive Queries, *International Journal of Computers and Their Applications*, Vol. 20, No. 1, pages 54-63, March, 2013

2. P. Patel, C. Langin, **F. Yu**, S. Rahimi, Network Intrusion Detection Types and Computation, *International Journal of Computer Science and Information Security*, Vol. 10, No. 1, pages 14-21, January, 2012

1. **F. Yu**, D. Oyana, W.-C. Hou, M. Wainer, Approximate Clustering on Data Streams Using Discrete Cosine Transform, *The Journal of Information Processing System*, Vol.6, No.1, pages 67-78, 2009

3.2 Peer-Reviewed Conference Publications

33. X.-J. Min, T. Kasamias, M. Wagner, A. Ogunbayi, **F. Yu**, Identification and Analysis of Alternative Splicing in Soybean Plants, *2022 International Conference on Bioinformatics and Computational Biology (BI-COB'22)*, 83, pp. 1-9, 2022, DOI: 10.29007/2qv6

32. S. Cal, E. Cheng, **F. Yu**, Optimized Bootstrap Sampling for σ-AQP Error Estimation: A Pilot Study, *30th International Conference on Software Engineering and Data Engineering (SEDE'21)*, 2021, pp. 144–153. DOI: 10.29007/bkw9

31. **F. Yu**, W.-C. Hou, CS*: Approximate Query Processing on Big Data using Scalable Join Correlated Sample Synopses, *2019 IEEE International Conference on Big Data (IEEE BigData 2019)*, December 9-12, 2019, Los Angeles, CA, USA, pp. 583-592

30. D.S. Wilson, W.-C. Hou, **F. Yu**, Scalable Correlated Sampling for Join Query Estimations on Big Data, *28th International Conference on Software Engineering and Data Engineering (SEDE'19)*, September 30 - October 2, 2019, San Diego, CA, USA, pp. 1-10, DOI: 10.29007/87vt (**Best Paper Award**)

29. M. Hamdi, **F. Yu**, and W.-C. Hou, A Denormalization Approach to Answering Join Queries, *The Tenth International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2019)*, June 2-6, Athens, Greece, pp. 17-27

28. R. Almutairi, M. Hamdi, **F. Yu**, W.-C. Hou, Performance Evaluations of Two Fast Join Query Processing Methods: Join Core and Join Indices, *The 33rd International Conference on Computers and Their Applications (CATA 2018)*, March 19-21, 2018, Las Vegas, Nevada, USA, pp. 104-109

27. G. Peng, **F. Yu**, Y. Peng, Co-membership, Networks Ties, and OSS Success: An Investigation Controlling Alternative Mechanisms for Knowledge Flow, *The 51th annual Hawaii International Conference on System Sciences (HICSS 2018)*, January 2018, pp. 4252-4260

26. M. Hamdi, **F. Yu**, S. Alswedani, W.-C. Hou, Efficient Data Structure for Fast Join Query Processing, *IEEE Technically Sponsored Future Technologies Conference (FTC) 2017*, November 29-30, 2017, Vancouver, BC, Canada, pp. 483-492

25. M. Hamdi, **F. Yu**, S. Alswedani, W.-C. Hou, Storing Join Relationships for Fast Join Query Processing, *28th International Conference on Database and Expert Systems, Applications (DEXA 2017)*, August 28 - 31, 2017, Lyon, Francem, pp. 167-177

24. M. Hamdi, **F. Yu**, W.-C. Hou, Fast Processing of Join Queries with Instant Response, *Computing Conference 2017*, July 18-20, 2017, London, United Kingdom, pp. 352-362

23. **F. Yu**, B. J. Latronica, T. J. Matacic, E. S. Jones, OB-Tree: Accelerating Data Cleaning in Out-of-Core

Column-Store Databases, *IEEE International Congress of Big Data 2017*, June 25 - June 30, 2017, Honolulu, Hawaii, USA, pp. 185-192

22. W. Xiong, M. Hamdi, **F. Yu**, and W.-C. Hou, A Hierarchical Precedence Concurrency Control Protocol for High Data Contention Database Environments, *ISCA 29th International Conference on Computer Applications in Industry and Engineering (CAINE-2016)*, September 26-28, 2016, Marriott Tech Center, Denver, USA, pp. 29-35

21. **F. Yu**, T. Matacic, W. Xiong, M. Hamdi, W.-C. Hou, Data Cleaning in Out-of-Core Column-Store Databases: An Index-Based Approach, *The 15th International Conference on Information & Knowledge Engineering*, July 25-28, 2016, Las Vagas, Nevada, USA, pp. 16-22

20. M. Hamdi, W. Xiong, **F. Yu**, W.-C. Hou, A Prudent-Precedence Concurrency Control Protocol for High Data Contention Main Memory Databases, *The 15th International Conference on Information & Knowledge Engineering*, July 25-28, 2016, Las Vagas, Nevada, USA, pp. 3-9

19. **F. Yu**, E. S. Jones, Hastening Data Retrieval on Out-of-Core Column-Store Databases using Offset B+-Tree, *28th International Conference on Computer Applications in Industry and Engineering (CAINE 2015)*, October 12-14, 2015, Hilton San Diego/Harbor Island, San Diego, California, USA, pp. 313-318

18. **F. Yu**, W.-C. Hou, A Framework of Write Optimization on Read-Optimized Out-of-Core Column-Store Databases, *26th International Conference on Database and Expert Systems Applications (DEXA 2015)*, Valencia, Spain, September 1-4, 2015, pp. 155-169

17. **F. Yu**, E. S. Jones, W.-C. Hou, Write Optimization using Asynchronous Update on Out-of-Core Column-Store Databases in Map-Reduce, *IEEE International Congress on Big Data*, June 27 - July 2, 2015, New York, USA, pp. 720-723

16. **F. Yu**, C. Luo, W.-C. Hou, E. S. Jones, Online Data Cleaning for Out-Of-Core Column-Store Databases with Timestamped Binary Association Tables, *30th International Conference On Computers And Their Applications (CATA 2015)*, Honolulu, Hawaii, USA, March 9-11, 2015, pp. 407-412

15. K. Neizer-Ashun, **F. Yu**, J. Meinken, X.-J. Min and G.-H. Chang, Prediction of Plant Protein Subcellular Locations, *7th International Conference on Bioinformatics and Computational Biology (BICOB 2015)*, Honolulu, Hawaii, USA, March 9-11, 2015, pp. 91-96.

14. **F. Yu**, C. Luo, W.-C. Hou, E. S. Jones, Asynchronous Update on Out-of-Core Column-Store Databases Utilizing the Time stamped Binary Association Table, *27th International Conference on. Computer Applications in Industry and Engineering (CAINE 2014)*, New Orleans, Louisiana, LA, October 13-15, 2014, pp. 215-220.

13. C. Luo, **F. Yu**, W.-C. Hou, S. He, Containment Join in the Presence of Nested Data Nodes, *27th International Conference on. Computer Applications in Industry and Engineering (CAINE 2014)*, New Orleans, Louisiana, LA, October 13-15, 2014, pp. 221-226.

12. **F. Yu**, W.-C. Hou, C. Luo, Sample Trace: Deriving Fast Approximation for Repetitive Queries, *6th International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2014)*, April 20 - 24, 2014, Chamonix, France, pp.59-65

11. **F. Yu**, W.-C. Hou, C. Luo, Histogram and Sample Revisited: Estimation Error Analysis For Multidimensional Range Selection Queries, *29th International Conference on Computers and Their Applications (CATA 2014)*, Las Vegas, NV, March 24-26, 2014, pp. 79-84

10. **F. Yu**, W.-C. Hou, C. Luo, D. Che, M. Zhu, CS2: A New Database Synopsis for Query Estimation, *2013 ACM SIGMOD International Conference on Management of Data (SIGMOD 2013)*, New York, New York, June 22-27, 2013, pp.469-480

9. **F. Yu**, C. Luo, W.-C. Hou, M. Wainer, A Prudent-precedence Concurrency Control Protocol for High Data Contention Database Environments, *28th International Conference on Computers and Their Applications (CATA 2013)*, Sheraton at Waikiki, Honolulu, Hawaii, USA, March 4-6, 2013, pp.289-294

8. C. Luo, W.-C. Hou, S. He, **F. Yu**, Using Cached Results to Expedite XML Query Evaluations, *25th International Conference on Computer Applications in Industry and Engineering (CAINE 2012)*, New Orleans, Louisiana, USA, November 14-16, 2012, pp. 81-85

7. **F. Yu**, W.-C. Hou, C. Luo, A Framework for Re-optimizing Repetitive Queries, *International Conference on Information and Knowledge Engineering (IKE 2012)*, Las Vegas, Nevada, USA, July 16-19, 2012, pp. 72-78

6. **F. Yu**, W.-C. Hou, M. Wainer, C. Luo, D. Che, Sufficient Statistics for Re-optimization Repetitive Queries, *27th International Conference on Computers and Their Applications (CATA 2012)*, Las Vegas, Nevada, USA, March 12-14, 2012, pp.282-287

5. **F. Yu**, W.-C. Hou, C. Luo, Q. Zhu, D. Che, Join Selectivity Re-estimation for Repetitive queries in Databases, *22nd DEXA Conference, Toulouse (DEXA 2011)*, France, Aug. 29 - Sept. 2, 2011, pp.420-427

4. C. Luo, **F. Yu**, W. Hou, Z. Jiang, D. Che, S. He, IRTA: An Improved Threshold Algorithm for Reverse Top-K Queries, *13th International Conference on Enterprise Information Systems (ICEIS 2011)*, Beijing, China, Jun 8- 11, 2011, pp.135-140

3. **F. Yu**, C. Luo, W.-C. Hou, D. Che, Selectivity Re-estimation for Repetitive Queries in Databases, *International Conference on Enterprise Information Systems and Web Technologies (EISWT 2010)*, Florida, USA, July 12-14, 2010, pp.12-18

2. J. Zhang, W.-C. Hou, **F. Yu**, D. Che, Enhancing SQL with Set-Comparison Operators, *24th International Conference on Computers and Their Applications (CATA 2009)*, April 8-10, 2009, New Orleans, Louisianan, USA, pp.51-56

1. C. Luo, Z. Jiang, W.-C. Hou, **F. Yu**, Q. Zhu, A Sampling Approach for XML Query Selectivity Estimation, *12th International Conference on Extending Database Technology (EDBT 2009)*, March 23-26, 2009, Saint Petersburg, Russia, pp.335-344

3.3 Book Chapter

1. **F. Yu**, C. Luo, X-G. Yu, W.-C. Hou, C.-F. Wang, M. Wainer, Range-Sum Queries over High Dimensional Data Cubes Using a Dynamic Grid File, Theoretical and Practical Advances in Information Systems Development: Emerging Trends and Approaches, Chapter 14, 2011

4 Funding and Awards

4.1 Extramural Funding Awarded (Total: $85,349)

1. Collaborative Research Experience for Undergraduates (CREU) 2018-19. Sponsor: Computing Research Association (CRA) and NSF, Title: Recommender Systems for Deal Websites. Co-PI. $13,500 awarded.
2. Amazon Web Service Research Education Grant 2017, Title: Develop Fast Data Cleaning on Large-Volume Columnar Data. PI. $8,500 awarded. ($6,099.45 extended in 2018.)
3. Amazon Web Service Research Education Grant 2016, Title: Improve Write Speed in Column-Store Databases using Timestamped Binary Association Table. PI. $5,000 awarded. ($1,300 extended in 2017.)
4. Amazon Web Service Research Education Grant 2016, Title: Large Scale TEM Image Feature Selection Study in a Cloud Computing Environment. Co-PI. Other Personnel: Zhang, Y. (PI, CSIS). $5,000 awarded.
5. Amazon Web Service Research Education Grant 2015, Title: Deriving Fast and Accurate Result Approximation of Big Query on a Big Data Platform. PI. $45,950 awarded.

4.2 Hornors and Awards

1. *Best Paper Award*, 28th International Conference on Software Engineering and Data Engineering (SEDE'19), 2019, San Diego, CA, USA
2. *Research Professorship*, Youngstown State University, 2015-2016, 2016-2017, 2018-2019, 2019-2020
3. *Distinguished Professor in Scholarship*, Youngstown State University, 2020
4. *Featured Reviewer*, ACM Computing Reviews, October, 2016

5 Academic Service

5.1 Editorial Board
1. Associate Editor, Circulation in Computer Science (ISSN: 2456-3692).
2. Associate Editor, International Journal of Database Management Systems (ISSN: 0975-5985)
3. Associate Editor, Transactions on Large-Scale Data and Knowledge-Centered Systems (Springer, ISSN: 1869-1994)

5.2 Scholarly Conference Service

5.2.1 Program Committee
1. IEEE International Conference on Big Data (IEEE BigData), 2021-current
2. IEEE International Conference on Machine Learning and Application, 2021
3. International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA), 2015-current
4. International Conference on Computer Applications in Industry and Engineering (CAINE), 2017-current
5. International Conference on Data Mining and Database (DMDB), 2020
6. International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA), 2015-current
7. International Conference on Human and Social Analytics (HUSO), 2017-current
8. International Conference on Software and Computer Applications (ICSCA), 2018-current
9. International Conference on Software Engineering and Data Engineering (SEDE), 2020-current
10. International Conference on Scientific and Statistical Database Management (SSDBM), 2021

5.2.2 Session Chair
1. IEEE International Conference of Big Data. Los Angeles, CA, December 9-12, 2019
2. 30th International Conference on Computers and Their Applications. Honolulu, Hawaii, March 9-11, 2015.
3. 29th International Conference on Computers and Their Applications. Las Vegas, NV, March 24, 2014.

5.3 Reviewer and Scholarly Judge

5.3.1 Peer-Reviewed Journals
1. ACM Transactions on Database Systems
2. ACM Computing Reviews
3. Circulation in Computer Science
4. Data and Knowledge Engineering
5. Database: The Journal of Biological Databases and Curation
6. International Journal for Computers and Their Applications
7. International Journal of Computer and Information Technology
8. International Journal of Data Mining, Modeling and Management
9. Information Sciences
10. Journal of Information Processing Systems
11. Open Life Sciences
12. Mathematical and Computational Applications
13. MDPI Sustainability

5.3.2 Peer-Reviewed Conferences
1. International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA)
2. International Conference on Computer Applications in Industry and Engineering (CAINE)
3. International Conference on Computers and Their Applications (CATA)
4. International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA)
5. International Conference on Human and Social Analytics (HUSO)
6. International Conference on Scientific and Statistical Database Management (SSDBM)
7. International Conference on Software and Computer Applications (ICSCA)

6 University Service

6.1 Departmental Assignment

1. Graduate Committee (Member), 2014–current
2. Honors and Scholarship Committee (**Chair**), 2015–2021
3. Library Committee (**Chair**), 2014–2019
4. Public Relation Committee (Member), 2015–2021
5. Resource Planning Committee (Member), 2015–2021
6. Faculty Search Committee (Member), 2013–2014, 2017–2018

6.2 College and University Assignment

1. Graduate College Admission Appeals Committee, 2021–current
2. Security and Policy Advisory Council (SPAC), 2018–2021
3. Program Review Committee of the Doctor of Education at YSU, 2017
4. Information Technologies Advisory Committee (ITAC), 2015–2018
5. Academic Senator (Representing CSIS Department), 2015–2016

7 Teaching

7.1 Undergraduate Courses

1. CSCI 3710 Introduction to Discrete Structures
2. CSIS 3722 Development of Databases
3. DATX 5801 Data Management (Designed for Data Analytics Certificate Program)
4. CSIS 4890 Computer Project (Senior Project)

7.2 Graduate and Swing Courses

1. CSCI 6950 Advanced Database Systems
2. CSCI 6951 Data Mining
3. CSCI 6971 Big Data and Cloud Computing (Formerly: CSCI 5895P Special Topics in Cloud Computing)
4. CSCI 6990 Computer Science Project (Master Project)
5. CSCI 6996 Independent Study
6. CSCI 6999 Thesis

7.3 Course Prepared

1. CSCI 4819 Parallel Computing
2. CSCI 5801 Software Engineering

8 Student Cultivation

8.1 Master Thesis Advised

1. Semih Cal (2021)
2. David Wilson (2018)
3. Tyler J. Matacic (2016)
4. Eric S. Jones (2015)
5. Nawal N Moatassem (2015)

8.2 Master Thesis Committee Member

1. Bhavya Sree Yadala (2021)
2. Ahmet Kapkic (2021)
3. Tabitha Kemboi (Math, 2021)
4. Joseph Costell (2020)
5. Burak Cetin (2020)
6. Astha Syal (2019)

7. Sahaj Bhattarai (2019)
8. Nishita Yedla (2018)
9. Brian Powell (2016)
10. Sai Krishnanand Nagavelli (2016)
11. Paranjith Lohiya (2015)
12. Ahraz Husain (2015)
13. Leela krishna Yenigalla (2014)

8.3 Master Project Advised

1. Marian Ansah (2020)
2. Prakyath Kumar Reddy Vallapu Reddy (2020)
3. Sai Teja Tirunamala (2020)
4. Sumanth Chowdary Uppala (2019)
5. Arkish Shah (2019)
6. Todd Jones (2019)
7. Brandon Latronica (2018)
8. Swetha Yelamanchi (2018)
9. Tasha Wells (2018)
10. Rohith Kumar Juvvadi (2018)
11. Apuroop Naidu Sadhanala (2017)
12. Naveen thumma (2016)
13. Avinash Naidu Suvvari (2015)
14. Priyanka Bheri (2015)
15. Ravi Kumar Chunchu (2015)
16. Wilfred C Mutale (2014)