YEONSU PARK

+82 10-9688-7293 Pohang, Republic of Korea
yspark@dblab.postech.ac.kr linkedin.com/in/yeonsu-park yspark-dblab.github.io

SUMMARY

I am a postdoctoral research scientist in Data Systems Laboratory at POSTECH, under the supervision of Professor Wook-Shin Han. My research interests include big data processing, query processing, and query optimization. Recently, I have been committed to developing new techniques to improve the ability of Apache Spark to process a massive number of small queries efficiently.

EDUCATION

Ph.D. in Computer Science and Engineering, POSTECH — Feb. 2018 - Feb. 2024
Advisor: Prof. Wook-Shin Han

B.S. in Software Engineering, Sungkyunkwan University — Mar. 2011 - Feb. 2017
Graduated with 1st rank in Dept. of Software
GPA: 4.35/4.5 (Major-only GPA: 4.43/4.5)
Took leave of absence for two years (for mandatory military service)

EMPLOYMENT

Postdoctoral Research Scientist, POSTECH, Republic of Korea — Feb. 2024 - Present
Researcher, POSTECH, Republic of Korea — Oct. 2017 - Feb. 2018
Software Engineer Intern, NCSOFT, Republic of Korea — Jan. 2012 - Feb. 2012

RESEARCH INTERESTS

Big Data Processing, Database Query Processing and Optimization, Algorithms

PUBLICATIONS

Peer-reviewed Conference Papers

[1] QaaD (Query-as-a-Data): Scalable Execution of Massive Number of Small Queries in Spark
Yeonsu Park, Byungchul Tak, and Wook-Shin Han
ACM SIGMOD 2023 (Top Database Conference)

[2] G-CARE: A Framework for Performance Benchmarking of Cardinality Estimation Techniques for Subgraph Matching
Yeonsu Park, Seongyun Ko, Sourav S. Bhowmick, Kyoungmin Kim, Kijae Hong, and Wook-Shin Han
ACM SIGMOD 2020 (Top Database Conference)

[3] A Survey on Worst-case Optimal Join Algorithms
Yeonsu Park, Taesung Lee, Seung-Min Lee, Junseung Hwang, and Wook-Shin Han
Korean Information Science Society Conference, 2018

[4] A Survey of Methods for Dynamic Graph Updates on the State-of-the-art Graph Processing Systems
Seung-Min Lee, Jeong-Hwan Kim, Byeonghoon So, Yeonsu Park, and Wook-Shin Han
Korean Information Science Society Conference, 2018

[5] Performance Evaluation of RocksDB Depending on Sync Option
Yeonsu Park, Gihwan Oh, Jong-baek Lee, Woon-Hak Kang, and Sang-Won Lee
Korean Information Science Society Conference, 2014

Dissertation

[6] Scalable Execution of Massive Number of Small Queries in Spark
Yeonsu Park
Ph.D. Dissertation, 2024

Patents

[7] DISTRIBUTED PROCESSING SYSTEM AND METHOD FOR PROCESSING DATA
Wook-Shin Han, Yeonsu Park, and Kijae Hong
KR Patent No. 10-2022-0110236, 2022

[8] ELECTRONIC APPARATUS AND DATA PROCESSING METHOD THEREOF, AND SYSTEM FOR DISTRIBUTED PROCESSING
Young Hwa Lee, Wook-Shin Han, Hyeonji Kim, and Yeonsu Park
KR Patent No. 10-2021-0172678, 2021

AWARDS AND HONORS

Google Conference Scholarship 2023
Graduation with 1st rank in Dept. of Software, Sungkyunkwan University 2017
ACM International Collegiate Programming Contest (ACM-ICPC) World Finals 2014
- *Special Award*
- *45th Place*

ACM International Collegiate Programming Contest (ACM-ICPC) Asia Regional (Korea Site) 2013
- *4th Place*

ACM International Collegiate Programming Contest (ACM-ICPC) World Finals 2013
- *48th Place*

ACM International Collegiate Programming Contest (ACM-ICPC) Asia Regional (Korea Site) 2012
- *2nd Place*

Sungkyun Software Scholarship 2011 - 2016
Dean's List for seven semesters, College of Computing, Sungkyunkwan University 2011 - 2016
Korea Olympiad in Informatics (KOI) 2009
- *Silver Medal*

PROJECTS

Learning to Construct Cost-efficient Batches of Small Queries in Spark 2023 - Present
Ph.D. Student & Postdoctoral Research Scientist, POSTECH Pohang, Republic of Korea

- Achieved 3.4× performance speed-ups compared to the state-of-the-art technique in Spark for a massive number of small queries by constructing cost-efficient batches of queries.
- Planning to submit at a top-tier database conference.

Scalable Execution of Massive Number of Small Queries in Spark 2022 - 2023
Ph.D. Student, POSTECH Pohang, Republic of Korea

- Achieved substantial performance improvement in Spark for small query workloads by proposing and implementing a query merge-based technique, resulting in 10.6× to 36.6× faster processing compared to standard Spark executions.
- Published at SIGMOD 2023.

Scalable Sequential Pattern Mining in Spark 2020 - 2022
Ph.D. Student, POSTECH (collaborated with Samsung Electronics) Pohang, Republic of Korea

- Parallelized the cSPADE algorithm in Spark, achieving a 100× improvement in scalability compared to the sequential pattern mining algorithm of Spark MLlib.

Performance Benchmarking of Cardinality Estimation Techniques for Subgraph Matching 2018 - 2020
Ph.D. Student, POSTECH Pohang, Republic of Korea

- Proposed and developed a comprehensive framework for cardinality estimation techniques, enabling the realization of existing methods and providing insights on their performance, by identifying serious accuracy issues in various scenarios and datasets.
- Discovered that a simple method designed for relational data consistently outperforms all others on graph data.
- Published at SIGMOD 2020.

ACADEMIC TALKS

QaaD (Query-as-a-Data): Scalable Execution of Massive Number of Small Queries in Spark

- ACM SIGMOD 2023, Seattle, WA, USA Jun. 2023

G-CARE: A Framework for Performance Benchmarking of Cardinality Estimation Techniques for Subgraph Matching

- Top Conference Session, Korea Computer Congress 2020 (Virtual) Jul. 2020
- ACM SIGMOD 2020, Portland, OR, USA (Virtual) Jun. 2020
- SAP Labs Korea, Seoul, Republic of Korea Nov. 2019

TEACHING EXPERIENCE

Teaching Assistant, CSED421: Database System, POSTECH, Pohang, Republic of Korea Spring 2021
Teaching Assistant, Advanced Data Programming, Samsung Electronics, Online 2020
Teaching Assistant, CSED421: Database System, POSTECH, Pohang, Republic of Korea Fall 2020
Teaching Assistant, CSED421: Database System, POSTECH, Pohang, Republic of Korea Fall 2019

SKILLS

Programming Languages C/C++, Python, Scala, Bash
Software & Technologies Big Data Framework (Apache Spark), Databases