

Wonbae Kim

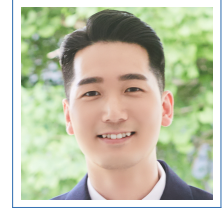
Curriculum Vitae

106-dong 605, UNIST-gil 50
44919 Ulsan
Korea

☎ +82 10 5522 5419

✉ wonbaekimys@gmail.com

🌐 <https://wonbaekimys.github.io>



Passionate about learning new things and developing innovative technologies.

Research Interests

- Big Data Systems
- Database Systems
- Distributed Systems
- Parallel Computing
- Machine Learning & Deep Learning
- Persistent Memory

Education

- Mar 2017 – Feb 2023 (Expected) **Ph.D. in Computer Science and Engineering**, *Ulsan National Institute of Science and Technology (UNIST)*, Ulsan, Korea, (Co-advisor: Prof. Young-ri Choi and Prof. Beomseok Nam)
- February 2017 **Master of Engineering in Computer Science**, *Ulsan National Institute of Science and Technology (UNIST)*, Ulsan, Korea
- February 2014 **Bachelor's degree in Computer Science and Engineering**, *Ulsan National Institute of Science and Technology (UNIST)*, Ulsan, Korea

Publications

- July 2022 **Wonbae Kim**, Chanyeol Park, Dongui Kim, Hyeongjun Park, Young-ri Choi, Alan Sussman, Beomseok Nam. "ListDB: Union of Write-Ahead Logs and SkipLists for Incremental Checkpointing on Persistent Memory." In *16th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2022)*, Carlsbad, CA, USA
- May 2022 Sunghwan Ahn, Hyeongjun Park, Vicente Adolfo Bolea Sanchez, Deukyeon Hwang, **Wonbae Kim**, Alan Sussman, and Beomseok Nam. "VeloxDFS: Streaming Access to Distributed Datasets to Reduce Disk Seeks." In *22nd IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGrid 2022)*, Taormina, Italy
- April 2022 Jonghyeon Yoo, Hokeun Cha, **Wonbae Kim**, Wook-Hee Kim, Sung-Soon Park, and Beomseok Nam. "Pivotal B+tree for Byte-Addressable Persistent Memory." *IEEE Access* 10
- October 2020 Soojeong Cho, **Wonbae Kim**, Sehyeon Oh, Changdae Kim, Kwangwon Koh, and Beomseok Nam. "Failure-Atomic Byte-Addressable R-tree for Persistent Memory." *IEEE Transactions on Parallel and Distributed Systems* 32

- September 2017 Vicente Adolfo Bolea Sanchez, **Wonbae Kim**, Youngmoon Eom, Kibeom Jin, Moohyeon Nam, Deukyeon Hwang, Jik-Soo Kim, and Beomseok Nam. "EclipseMR: Distributed and Parallel Task Processing with Consistent Hashing." In *IEEE International Conference on Cluster Computing (CLUSTER 2017)*, Honolulu, Hawaii, USA
- June 2017 **Wonbae Kim**, Young-ri Choi, and Beomseok Nam. "Coalescing HDFS Blocks to Avoid Recurring YARN Container Overhead." In *IEEE 10th International Conference on Cloud Computing (CLOUD 2017)*, Honolulu, Hawaii, USA
- May 2017 **Wonbae Kim**, Young-ri Choi and Beomseok Nam. "Mitigating YARN Container Overhead with Input Splits." In *17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2017)*, Madrid, Spain

Experience

- Jan 2018 – Jun 2018 **Intern**, *BrightEdge*, Foster City, CA, USA
- Spring 2013 – Present **Research Assistant**, *Data Intensive Computing Lab. (DICL)*, SKKU, Suwon, Korea
- Worked on a distributed semantic caching framework for MapReduce
 - Worked on reducing container initialization overhead of Hadoop YARN
 - Worked on improving the performance of distributed runtime of Tensorflow
 - Working on exploiting persistent memory for index structures and key-value stores.
- Dec 2015 – Feb 2016 **Visiting Student (Prof. Ikhlal Sidhu)**, *University of California Berkeley*, Berkeley, CA, USA
- Spring 2014 **Teaching Assistant**, *Engineering Programming Course*, UNIST, Ulsan, Korea
- Basic C++ programming course for freshmen.

Research Projects

- 2019 – present **Main Contributor**, *ListDB Key-Value Store*, <https://github.com/DICL/listdb>
- ListDB is an LSM tree-based key-value store optimized for persistent memory. As the first author, I contributed most of the design and implementation of ListDB. ListDB's paper was published at OSDI '22.
- 2016 – 2020 **Student Researcher**, *National Project for High-Performance Distributed Deep Learning*
- This project is a nation-wide project that many research labs in Korea participate in. This project covers various fields ranging from machine learning applications to system-level research for distributed model training. I worked on an asynchronous decentralized approach using RDMA to solve the network communication overhead of distributed training in Tensorflow.
- National R&D achievements: SW registration number C-2021-021832
- 2014 – 2021 **Contributor**, *Velox Bigdata Framework*, <https://github.com/DICL/VeloxMR>
- Velox is a decentralized big data processing framework implemented in C++, consisting of distributed file systems, in-memory caches, MapReduce and SQL engine. I worked on implementing MapReduce engine.

Honors and Awards

- July 2022 **Student Grant Award - OSDI 2022**
- June 2017 **Best Student Paper Award - IEEE CLOUD 2017**

Technical and Personal Skills

- **Programming Languages:** C, C++, Java, Python, Shell
- **System Software/Frameworks/Libraries:** Linux, RocksDB, Tensorflow, Hadoop, Spark, CUDA, MPI, RDMA, GRPC, Protocol Buffers, PMDK
- **Other:** git, Tex, vim, gnuplot, gcc, CMake, Maven, bazel

References

- **Young-ri Choi**
Professor
Department of Computer Science and Engineering
Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea
E-mail: ychoi@unist.ac.kr
- **Beomseok Nam**
Associate Professor
College of Computing and Informatics
SungKyunKwan University (SKKU), Suwon, Korea
E-mail: bnam@skku.edu
- **Sam H. Noh**
Professor
Graduate School of Artificial Intelligence
Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea
E-mail: samhnoh@unist.ac.kr
- **Woongki Baek**
Associate Professor
Department of Computer Science and Engineering
Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea
E-mail: wbaek@unist.ac.kr