

HEERAK LIM / 임희락

1992.12.29 / Seoul, Korea.

+82 10-4814-3800 / lrocky1229@gmail.com

Summary

With 4+ years of experience working on distributed database systems and storage systems in a development team and academia. I look forward to bringing my extensive technical and analytical skills to a development position in system, storage and database

Work Experience

CSV Import performance optimization with SIMD instructions and concurrency-aware data structure.

SAP Labs Korea, ImportExport team (Aug 2019 -)

- Designed and implemented a parallel CSV parsing with SIMD instructions
- Designed and implemented a concurrency-aware-queue with limited queue size for memory-efficient CSV data import
- Designed and implemented a pipelined architecture for high-performance CSV data import
- Reduced turn-around time performance for CSV data import up to 52% and reduced memory peak up to 56%

Parquet format introduction to SAP HANA Database

SAP Labs Korea, ImportExport team (Aug 2019 -)

- Designed and Implemented a data Import/Export features to support Apache Parquet, an open source data storage format, for SAP HANA database
- It reduced the storage footprint of a data export feature about 74% compared with CSV format and consequently reduced storage and network costs of SAP HANA on cloud service.

Performance optimization for Parquet format with row group size estimation

SAP Labs Korea, ImportExport team (Aug 2019 -)

- Design and implementation for a row group size estimation of Parquet data export feature to increase parallel I/O request
- Made SAP HANA database leverages the whole concurrency resource to export data into Parquet format regardless of the distribution or type of data.
- Elapsed time for data export as Parquet format from SAP HANA is reduced up to 73%

Development Radio Tomographic Imaging (RTI) system on low-power sensor networks

Commonwealth Scientific and Industrial Research Organisation (CSIRO) Australia. Data61, Distributed Sensing System Team(Research Intern). (Dec 2016 – Feb 2017)

- Designed and implemented multiple sensor network nodes synchronization algorithm
- Designed and Implemented RTI application on Contiki OS
- Optimized network device driver to achieve superlative frequency
- Developed Image reconstruction utility using python

Academic Experience

LSM-tree based Database System Optimization using Application-Driven Flash Management

Distributed Computing System Laboratory (DCSLAB), Seoul National University (Sep 2018 –)

- Optimized the LSM-tree based Key-Value store database system by processing I/O considering append-only application context
- Modified RocksDB storage backend to manage Flash based storage device directly using user-space I/O library
- Read/Write throughput is increased by 20% and read latency is reduced by 43%
- Open-source contribution (bug fix)
- Contributed to international conferences and presented academic conferences(HiPC 2019, Middleware 2018)

Education

Master of Science in Computer Science

Seoul National University, Distributed Computing System Laboratory (DCSLAB) (2019),

Graduated with a bachelor's degree in Department of Software Engineering

Ajou University (2017)

Publications(International)

- **Lim, Heerak**, et al. "IsoKV: An Isolation Scheme for Key-Value Stores by Exploiting Internal Parallelism in SSD." 2019 IEEE 26th International Conference on High Performance Computing, Data, and Analytics (HiPC). IEEE, 2019.
- **Lim, Heerak**. "Application-Driven Flash Management: LSM-tree based Database Optimization through Read/Write Isolation." Proceedings of the Doctoral Symposium of the 19th International Middleware Conference. ACM, 2018.

Technical Skills

- Extensive experience in C++, C Programming language
- Deep understanding in Linux operating system and storage system
- Extensive experience of problem-solving and debugging on Linux/Unix environment
- practical experience in Test-Driven Development(TDD) practice
- Practical experience in Python

Areas of interest

- Database, Storage system, Distributed system
- Solidity and Ethereum smart contract
- System performance optimization
- Dev tools and automation.

Others

- Research Blog - <https://generousrocky.github.io/>