Hyogi Sim

HPC Systems Engineer (*R&D Associate*), Technology Integration Group
National Center for Computational Sciences — *simh@ornl.gov*
Oak Ridge National Laboratory — +1-865-574-6167

Work Experience

HPC Systems Engineer, Oak Ridge National Laboratory	*Feb. 2017 – present*
Postmasters Research Assistant, Oak Ridge National Laboratory	*Mar. 2015 – Jan. 2017*
Summer Intern, Oak Ridge National Laboratory	*Jan. 2013 – Aug. 2013*
Graduate Research and Teaching Assistant, Virginia Tech	*Aug. 2011 – Dec. 2013, Sep. 2013 – Feb. 2015*
Senior Developer, Pitapat Mobile, S. Korea	*Jan. 2011 – June 2011*
Part-Time Programmer, Dept. of English Education, Hanyang University, S. Korea	*Aug. 2009, Feb. 2010*
Part-Time Programmer, Metabuild co., S. Korea	*Dec. 2008*
Senior Developer, Golfschool co., S. Korea	*2002 – 2003*
Military Conscription in Combat Police Force, S. Korea	*May 1999 - Jul. 2001*

Education

Ph.D. in Computer Science, Virginia Tech, VA — *— expected in 2022*
- Advisor: Dr. Ali R. Butt
- Thesis: Storage System Support for Data-Intensive Computing in Large-Scale HPC Systems *(working title)*

M.S. in Computer Science, Virginia Tech, VA — *— Dec. 2014*
- Advisor: Dr. Ali R. Butt
- Thesis: AnalyzeThis: An Analysis Workflow-Aware Storage System

M.S. in Electronics and Computer Engineering, Hanyang University, S. Korea — *— Feb. 2008*
- Thesis: A Study of Performance Impact of Merging Storage Layers on Flash-Based DBMS

B.S. in Urban Planning, Hanyang University, S. Korea — *— Feb. 2005*

Skills & Expertise

Skillful in Linux application/system/kernel programming, parallel/distributed file systems, database systems.
- Programming Languages: C, C++, Python, JAVA, PHP, C#, BASH
- File & Storage: FUSE, Linux VFS, Device Mapper, SCSI OSD-2, Linux TGT, exofs, ext3
- Parallel & Distributed File Systems: GlusterFS, Lustre, CephFS, GPFS, NFS, HDFS, HPSS
- Non-Volatile Memory & SSD: Linux MTD, Flash Translation Layer, NVMe
- Databases: MySQL, PostgreSQL, SQLite, MS SQLServer, IBM DB2, WiSS, HyperDex, ElasticSearch
- Parallel Programming: MPI, Pthread, OpenMP, Spark
- Parallel I/O Library: netCDF, HDF5, SCR
- I/O Benchmarks/Profiling: fio, mdtest, IOR, TPC-C, TPC-H, blktrace, Darshan
- Tools: gcc, gdb, cscope, ctags, autotools, git, svn, eclipse, glade, visual studio, latex, gnuplot
- Web Development: HTML, CSS, Django, Jekyll, PHP, Javascript, JQuery, ASP.NET
- Container: Docker, OpenShift
- Package Management: RPM, Portage (Gentoo Linux), Pkgsrc (NetBSD), Ports (FreeBSD)

Project Experience in File and Storage Systems

Scalable data infrastructure for science — *2019 – present, Oak Ridge National Lab.*
- Developing a scientific data repository that allows data sharing among scientists.
- Developing a web portal that interacts with the backend storage.
- Deploying the services with containers on the OpenShift cluster.
- *(Python, Django, PostgreSQL, OpenShift, Docker, Globus, DOI)*

Metadata indexing framework for HPC archival storage systems — *2018 – present, Oak Ridge National Lab.*
- Developing a software framework that periodically collects metadata from HPSS and builds data catalog.
- Developing a framework that extracts metadata from well-known document files without user intervention.
- Developing command-line utilities for system administrators and users.
- *(HPSS, IBM DB2, PostgreSQL, Apache Tika, XDR, cURL, ElasticSearch, Kafka)*

UnifyFS: A userspace checkpoint file system for HPC burst buffers *2017 – present, Oak Ridge National Lab.*

- Exascale Computing Project (ECP): Software technologies/Data and Visualization/ExaIO
- Developing an ephemeral, distributed file system for node-local burst buffers to facilitate checkpointing of HPC applications.
- *(FUSE, MPI, PMIx, MDHIM, LevelDB, Mercury, Margo, Argobots, Slurm, LSF)*

Large-scale file system snapshot and workload analysis *2016 – 2019, Oak Ridge National Lab.*

- Analyzing the snapshot data of Spider II (32 PB Lustre) file system.
- Analyzing the eight year worth workloads of 80 PB HPSS archival storage system.
- *(Lustre, HPSS, Spark, SparSQL, Parquet, Python, MySQL, SQLite)*

File system-integrated search and discovery services for HPC *2015 – 2017, Oak Ridge National Lab.*

- Developing a file system-integrated metadata indexing framework that supports taggings in GlusterFS and CephFS.
- *(GlusterFS, CephFS, Linux, SQLite)*

Programming framework for processing-in-memory architecture *2015 – 2016, Oak Ridge National Lab.*

- Developing a programming library (C/C++) and runtime environment for facilitating data analysis tasks with PIM devices.
- Developing a kernel device driver that emulates PIM devices in a NUMA architecture.
- *(NUMA, Pthread, Intel TBB, Linux Kernel)*

Analysis-aware storage system for high performance computing *2013 – 2015, Oak Ridge National Lab./Virginia Tech*

- Developing an active execution framework by extending the SCSI T10 OSD-2 specification.
- Extending the Linux exofs file system to support the extended OSD protocol.
- Developing a FUSE file system that manages the array of active OSD devices.
- *(Linux Kernel, SCSI T10 OSD-2 Protocol, Linux TGT, FUSE, SQLite)*

Hierarchical data management in media servers with hybrid storage architecture *2007 – 2009, Hanyang University*

- Developing a content popularity analyzer that identifies popular media contents in a commercial media server.
- Developing a light-weight file system for storage class memory devices (PRAM) in media cache servers.
- *(Linux Kernel, Windows Media Server)*

Development of a transactional record storage for NAND flash memory *2006 – 2008, Hanyang University*

- Developing a DBMS that directly manages a NAND flash memory via Linux MTD to manage EPG data in a set-top box.
- Developing a new PostgreSQL buffer manager that generates an optimized LBA sequence for NAND flash memory.
- Developing an interactive framework that identifies a FTL mapping schema for NAND flash memory.
- *(Linux Kernel, Linux MTD, Wisconsin Storage System, PostgreSQL, ARM-based embedded board)*

Other Project Experience

Development of a web-based social marketing game *2011, Pitapat Mobile, S. Korea*

- Developing a web application that features social quiz games and integrating it to the Facebook via Facebook API.
- *(C#, Javascript, JQuery, HTML, REST, ASP.NET, SQL Server, Amazon EC2)*

Development of an image transfer library for a self-driving vehicle *2008, Metabuild co., S. Korea*

- Developing a system library that transfers requested pixel data from the CMOS camera to a lane-recognition module.
- *(Linux, C, ARM-based embedded board)*

Development of an online survey website *Feb 2010, Aug 2009, Dept. of English Education, Hanyang University, S. Korea*

- Developing a web application that dynamically visualizes online survey results with graphs.
- *(Linux, PHP, HTML, Javascript, MySQL, GD, Apache, Microsoft Excel)*

Construction of a protein function database *2006 – 2008, Hanyang University*

- Developing a GUI environment for verifying protein interaction graphs that are extracted automatically from academic papers.
- *(Windows, SQLServer, Delphi)*

Development of an online lesson website *2002 – 2003, Golfschool co., S. Korea*

- Developing and maintaining a commercial company website with online sports lessons and an integrated shopping mall.
- *(Linux, Apache, PHP, MySQL, HTML, Javascript)*

Publications

- Awais Khan, Hyogi Sim, Sudharshan S. Vazhkudai, Ali R. Butt, Youngjae Kim, *An Analysis of System Balance and Architectural Trends Based on Top500 Supercomputers*, In Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region (HPC Asia '21), January 2021
- Awais Khan, Hyogi Sim, Sudharshan S. Vazhkudai, Jinsuk Ma, Myeong-Hoon Oh, Youngjae Kim, *Persistent Memory Object Storage and Indexing for Scientific Computing*, In Proceedings of the IEEE/ACM Workshop on Memory Centric High Performance Computing (MCHPC '20) in conjunction with SC, GA, November 2020

- Hyogi Sim, Awais Khan, Sudharshan S. Vazhkudai, *An Analysis of System Balance and Architectural Trends Based on Top500 Supercomputers*, Oak Ridge National Laboratory, ORNL/TM-2020/1561, Aug. 2020
- Hyogi Sim, Awais Khan, S Vazhkudai, S Lim, Ali R. Butt, Youngjae Kim, *An Integrated Indexing and Search Service for Distributed File Systems*, IEEE Transactions on Parallel and Distributed Systems (TPDS), vol. 31, no. 10, pp. 23752391, 2020.
- Ali Anwar, Yue Cheng, Hai Huang, Jingoo Han, Hyogi Sim, Dongyoon Lee, Fred Douglis, Ali R. Butt, *Customizable Scale-Out Key-Value Stores*, IEEE Transactions on Parallel and Distributed Systems (TPDS), vol. 31, no. 9, pp. 20812096, 2020.
- Hyogi Sim, Sudharshan S. Vazhkudai, *Profiling the Usage of an Extreme-Scale Archival Storage System*, Proceedings of the 27th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOSTS '19), Rennes, France, October 2019
- Hyogi Sim, Geoffroy R. Vallée, Youngjae Kim, Sudharshan S. Vazhkudai, Devesh Tiwari, Ali R. Butt, *An Analysis Workflow-Aware Storage System for Multi-Core Active Flash Arrays*, IEEE Transactions on Parallel and Distributed Systems (TPDS), vol. 30, no. 2, pp. 271285, 2019.
- Hyogi Sim, Arnab K. Paul, Eli Tilevich, Ali R. Butt, *CSLIM: Automated Extraction of IoT Functionalities from Legacy C Codebases*, Proceedings of the 8th International Workshop on Computing and Networking for IoT and Beyond (ComNet-IoT 19) in conjunction with ICDCN '19, Bangalore, India, pages 6, January 2019
- Ali Anwar, Yue Cheng, Hai Huang, Dongyoon Lee, Jingoo Han, Hyogi Sim, Fred Douglis, and Ali R. Butt, *BESPOKV: Application Tailored Scale-Out Key-Value Stores*, Proceedings of the 2018 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '18), Dallas, TX
- Sangkuen Lee, Hyogi Sim, Sudharshan S. Vazhkudai, *A Programmable Shared-Memory System for an Array of Processing-In-Memory Devices*, Cluster Computing: The Journal of Networks, Software Tools and Applications, 2018
- Feiyi Wang, Hyogi Sim, Cameron Harr, Sarp Oral, *Diving into Petascale Production File Systems through Large Scale Profiling and Analysis*, In Proceedings of the 2nd Joint International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems (PDSW-DISC '17) in conjunction with SC, Denver, CO
- Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, Geoffroy R. Vallée, Seung-Hwan Lim, Ali R. Butt, *TagIt: An Integrated Indexing and Search Service for File Systems*, Proceedings of the 2017 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '17), Denver, CO
- Seung-Hwan Lim, Hyogi Sim, Raghul Gunasekaran, Sudharshan S. Vazhkudai, *Scientific User Behavior and Data-Sharing Trends in a Petascale File System*, Proceedings of the 2017 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '17), Denver, CO
- Sangkuen Lee, Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, *AnalyzeThat: A Programmable Shared-Memory System for an Array of Processing-In-Memory Devices*, IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID '17), Madrid, Spain, May 2017
- Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, Geoffroy R. Vallée, Seung-Hwan Lim, Ali R. Butt, *TagIt: An Integrated Search and Discovery Service for Extreme-Scale File Systems*, Poster in the 2016 USENIX Annual Technical Conference (ATC '16), Denver, CO, June 2016
- Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, Devesh Tiwari, Ali Anwar, Ali R. Butt, Lavanya Ramakrishnan, *AnalyzeThis: An Analysis Workflow-Aware Storage System*, Proceedings of the 2015 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC '15), Austin, TX
- Hyogi Sim, Youngjae Kim, Sudharshan S. Vazhkudai, Devesh Tiwari, Ali Anwar, Ali R. Butt, Lavanya Ramakrishnan, *AnalyzeThis: An Analysis Workflow-Aware Storage System*, Poster in the 2015 USENIX Annual Technical Conference (ATC '15), Santa Clara, CA, July 2015
- Hyogi Sim, Hoyoung Jung, Sungmin Park, Sooyong Kang, Jaehyuk Cha, *Identifying the FTL Mapping Scheme for USB Flash Devices*, The 4th International Conference on Convergence Technology and Information Convergence (CTIC), Oct. 2009
- Sooyong Kang, Sungmin Park, Hoyoung Jung, Hyogi Sim, Jaehyuk Cha, *Performance Tradeoffs in Using NVRAM Write Buffer for Flash Memory-based Storage Devices*, IEEE Transactions on Computers, vol. 58, iss. 6 (Jun. 2009), pp. 744-758
- Hoyoung Jung, Hyogi Sim, Sungmin Park, Sooyong Kang, Jaehyuk Cha, *LRU-WSR: Integration of LRU and Writes Sequence Reordering for Flash Memory*, IEEE Transactions on Consumer Electronics, vol. 54, iss. 3 (Aug. 2008)
- Sungmin Park, Hoyoung Jung, Hyogi Sim, Sooyong Kang, Jaehyuk Cha, *Using Non-Volatile RAM as a Write Buffer for NAND Flash Memory-based Storage Devices*, 2008 IEEE International Symposium on Modeling, Analysis & Simulation of Computer & Telecommunication Systems, MASCOTS 2008, Sept 8-10, Baltimore, MD
- Sungmin Park, Hoyoung Jung, Hyogi Sim, Sooyong Kang, Jaehyuk Cha, *Write Buffer-aware Address Mapping for NAND Flash Memory Devices*, 2008 IEEE International Symposium on Modeling, Analysis & Simulation of Computer & Telecommunication Systems, MASCOTS 2008, Sept 8-10, Baltimore, MD
- Hoyoung Jung, Kyunghoon Yoon, Hyogi Sim, Sungmin Park, Sooyong Kang, Jaehyuk Cha, *LIRS-WSR: Integration of*

LIRS and Write Sequence Reordering for Flash Memory, The 2007 International Conference on Computational Science and Its Applications, ICCSA LNCS 2007, Aug. 29

Presentations

- Hyogi Sim, *Extracting Metadata from the ORNL HPSS Archive to Improve its Usability*, "Knowledge Is Power: Unleashing the Potential of Your Archives through Metadata" BoF in ACM/IEEE International Conferenece for High Performance Computing, Networking, Storage, and Analysis (SC), Denver, CO, November 2019
- Hyogi Sim, *Making a Peta-Scale Archival Storage System Searchable*, High Performance Storage Systems User Forum 2019 (HUF 2019), Indiana University, Bloomington, IN, October 2019
- Hyogi Sim, *Profiling the Usage of an Extreme-Scale Archival Storage System*, High Performance Storage Systems User Forum 2019 (HUF 2019), Indiana University, Bloomington, IN, October 2019

Professional Services

Program Committee Member:

- BenchCouncil International Symposium on Benchmarking, Measuring and Optimizing (Bench): 2018, 2019, 2020

Journal Reviewer:

- IEEE Transactions on Storage (TOS)
- IEEE Transactions on Parallel and Distributed Systems (TPDS)
- Journal of Parallel and Distributed Computing (JPDC)

Recognitions

- Analysis Restaurant: In November 2015, *DEIXIS online magazine* featured AnalyzeThis storage system as a monthly highlight, based on the SC '15 paper (*https://deixismagazine.org/2015/11/analysis-restaurant/*).

Teaching Experience

Computer Organization II, Virginia Tech, VA — *Spring 2012, Fall 2011 (Teaching assistant)*

File Structure, Hanyang University, S. Korea — *Spring 2009, Fall 2007 (Leading lab sessions)*

Data Structures, Hanyang University, S. Korea — *Spring 2007 (Leading lab sessions)*

Object Oriented Programming, Hanyang University, S. Korea — *Spring 2005 (Leading lab sessions)*

Awards & Scholarships

Graduate Research Assistantship, Virginia Tech, VA — *Fall 2011, 2012, 2013, Spring 2012, 2014*

BK21 Scholarship, Hanyang Univeristy, S. Korea — *2009*

Academic Record Scholarship, Hanyang Univeristy, S. Korea — *2005*