Kangwoo Choi

+82-010-8964-2718 kangwoo.choi.antonio@gmail.com linkedin.com/in/kangwoo-choi-antonio/

Keywords: ML4DB, DB4ML

Devise & Implement novel methods based on Machine-Learning to estimate SQL Execution Cost and Output Size when SQL optimizer decides which SQL Plan is Optimal to Execute on Runtime

Technical Skills:

Languages: Python3, PyTorch, Cython, Pandas, ScikitLearn, R, C++, C,
Developer Tools: VS Code, VIM, GitHub
Technologies/Frameworks: RDBMS, SQL, NoSQL, Linux(SuSE), GPFS(File System), Jenkins

Language:

Fluency in English and Native in Korean
TOEFL: RC/LC/SPK/WRT(29/30/21/25)

Work Experience

SAP Labs Korea — April 2012 - Present

Software Developer, Research & Innovation Team — **January 2017 - Present**

Developed Filter Cost Model for *SAP HANA*(RDBMS) SQL Optimizer based on Machine Learning(*XGBoost, LightGBM, ElasticNet, MSCN, FCNN, etc.*) with various Datasets – Real DBs(*DMV, IMDB, CustomerDB*), Benchmark DBs(*JOB, TPCH, TPCDS, JCCH*), and Synthetic DB(*ongoing*)

- Explored Data Analysis on Real DBs(DMV, IMDB), Benchmark DBs(TPCH, TPCDS, JCCH), and newly generated and customized Synthetic DB for General Availability
- Designed Feature Set/Implemented Codes to automatically generate 30K+ SQL Statements for building Training Datasets with varying SQL Data Type, Predicate Type, Distinct Value Count, etc.

Led co-research/development of Output Size Estimator for SQL Operators using Machine Learning and Statistics

- (*Functionality Improvement*) Expand SQL Operator Coverage in Estimator by devising new Distinct Count Value Estimator for GroupBy SQL Operators – One Patent Filed regarding New GroupBy Estimator
- Led to co-research with Postech as a main counterpart and Published One Survey Paper regarding Size Estimators based on State-Of-The-Art Deep Learning Approaches in *SIGMOD 2022*
- Surveyed and Evaluated Relevant Papers on various Approaches to estimating Output size of SQL Operator Execution such as *NeuroCard* & *NARU* based on Auto-Regressive Models, FLAT based on Statistics combined with novel Ideas, MSCN(*Multi-Set Convolutional Network*) compared to Size Estimator of *SAP HANA*(RDBMS) SQL Optimizer

Led co-research of Persistent Memory, CXL, 3DS TSV 128GB DRAM related Projects w/ Samsung

- (*New Functionality*) Design and Implement Test Framework using PMDK(Persistent Memory Development Kit) to evaluate New Memory Technology made by Samsung – Persistent Memory using DRAM and SSD(called NVDIMM-C). This led to reducing TCO(Total Cost Ownership) and Publishing of one Paper in *HPCA 2020*
- (*New Functionality & Performance Improvement*) Identified the optimal major workload to offload into Near-Memory for Acceleration of In-Memory RDBMS(*HANA*), Designed Experiment to measure Performance Gain, and Conducted Experiment on Prototype of Accelerator developed by Samsung – Published one short Paper in *EDBT 2020*
- (*New Functionality*) Designed, Implemented, and Delivered End-to-End Test Workloads using Capture and Replay Feature in *SAP HANA* for newly developed Commercial Memory Technology(*3DS TSV 128GB DRAM*) in Samsung
- (*New Functionality*) Evaluated Compression ratio and Latency of New Compression Logic Embedded Memory Implemented in Hardware Level using various Customer Scenarios and Benchmark Tests

Quality Specialist, Quality Team — **April 2012 – December 2016**

- Maintained BW Test Frameworks and weakly Conducted to ensure Performance of SAP HANA (RDBMS) in BW Scale-out(one Application Server + four-node DBMS) Systems
- Developed Dev Infrastructure – Setup Test Environment including Hardware for testing new Features and measuring Performance of SAP HANA and GPFS shared Storage Setup for about 250+ Internal Developers and Developed initial Reporting Service for management

SAMSUNG Electronics **January 2010 – March 2012**

Software Engineer, Network Division

Implemented and Maintained Call Recovery System of PDN Gateway in LTE core network

• This supported 10M end-users / one rack for High Availability and Durability by Implementing Fast Recovery Module

ROK of Air Force **January 2007 – December 2009**

First Lieutenant Officer, Civil Engineering

EDUCATION

Seoul National University (SNU) **March 2002 – August 2005**

B.S. in Electrical Engineering

Relevant Coursework

• Quantum Physics(*Dep. Physics*) • Microwave Theory(*grad course*) • Non-Linear Microwave Circuit Analysis and Design(*grad course*)

• Electromagnetic Field Theory • Linear Algebra • Complex Analysis • Partial Differential Equation

• Advanced Electromagnetic Field Theory

Korea National Open University (KNOU) **March 2020 – February 2022**

B.S in Computer Science *Graduated with High Honor Awards(4.44/4.5, within 3%)*

Relevant Coursework

• DBMS • C++ Programming • Data Structure • Algorithm

• Probability Concept and Applications • Compiler Construction • Programming Language

• Statistical Understanding of Deep Learning • Computer Architecture • Artificial Intelligence

• C Programming • Cloud Computing • Computer Graphics • Understanding of Big Data

Korea National Open University (KNOU) **March 2022 – Present**

B.S in Statistics and Data Science

Relevant Coursework

• Multi-Variate Analysis • Regression Model Analysis • Predictive Analytics • Data Visualization

AWARDS

Graduated with High Honors, KNOU **February 23, 2022**

Rockstar of SAP Labs Korea, SAP Labs Korea **February 2, 2021**

PUBLICATION

Conference

C1. *(Postech)*Kyoungmin Kim, Jisung Jeong, In Seo, Wook-Shin Han, *(SAP)***Kangwoo Choi**, Jaehyok Chong, "Learned Cardinality Estimation: An In-depth Study", Proceedings of the 2022 International Conference on Management of Data (***SIGMOD***), June 11, 2022

C2. *(SAP)*Donghun Lee, Minseon Ahn, Jungmin Kim, **Kangwoo Choi**, Oliver Rebholz, *(SAMSUNG)*Andrew Chang, Jongmin Gim, Jaemin Jung, Vincent Pham, Krishna Malladi, YangSeok Ki, **"Optimizing Data Movement with Near-Memory Acceleration of In-memory DBMS"**, International Conference on Extending Database Technology (***EDBT***), April, 2020

C3. *(SAMSUNG)*Changmin Lee, Wonjae Shin, Dae Jeong Kim, Yongjun Yu, Sung-Joon Kim, Taekyeong Ko, Deokho Seo, Jongmin Park, Kwanghee Lee, Seongho Choi, Namhyung Kim, Vishak G, Arun George, Vishwas V, *(SAP)*Donghun Lee, **Kangwoo Choi,** Changbin Song, Dohan Kim, Insu Choi, Ilgyu Jung, Yong Ho Song, Jinman Han, **"NVDIMM-C: A Byte-Addressable Non-Volatile Memory Module for Compatibility with Standard DDR Memory Interfaces"**, 2020 IEEE International Symposium on High Performance Computer Architecture (***HPCA***), Feb 1, 2020

White Papers

W1. Accelerating Analytical Banking Solutions with Large-scale System Memory A Proof of Concept by Lenovo, Samsung, and SAP, 2018/06, 58597enUS

Patents

P1. **Kangwoo Choi,** Daeun Lee, Dong Hun Lee, "**Group-by Cardinality Estimation using Data-driven Learning**", Patent Application 17/979643, filed 2-Nov-2022

P2. Dong Hun Lee, Jungmin Kim, **Kangwoo Choi**, Minseon Ahn, Oliver Rebholz, "**Optimizing Data Movement with Near-Memory Acceleration**", Patent Application 62/982683, filed 27-Feb-2020