Memory based AI & Data Analytics Solutions

Euicheol Lim, SK hynix

Abstract: The advent of the AI/Big data era is causing a rapid shift in computer architecture from computing-centric to data-centric. In order to provide a competitive AI-based service, it is necessary to process a huge amount of data quickly in the data center, and it is also needed to process more data at the edge server. However, in the traditional computing architecture where data movement occurs in the order of cache - memory - storage, data movement consumes more energy rather than data operation itself and the performance of the system largely depends on it. Near data processing can be a solution to address the data movement issue by offloading certain tasks near or in memory side, thus improving performance and energy efficiency. This course shows the demand of near data processing from the data pipeline and operation flow of the specific AI/Big data service system, and with a specific solution examples I'd like to explain what type of near data processing architecture is feasible. The analog computing in memory that processes data and processing in analog cell array will be introduced as an extreme example of near data processing. And as a more practical solution, the computational memory card solution will be covered as well. Also, we will discuss which part of the overall AI service system each solution will play a role in. In this course, we mainly focus on the solution and system perspectives rather than the device-level technology.

Euicheol Lim is a Research Fellow and leader of system architecture team in memory system research, SK hynix. He received the B.S. degree and the M.S. degree from Yonsei University, Seoul, Korea, in 1993 and 1995, and the Ph.D. degree from Sungkyunkwan University, suwon, Korea in 2006. Dr. Lim joined SK Hynix in 2016 as a system architect in memory system research. Before joining SK Hynix, he had been working as an SoC architect in Samsung Electronics and leading the architecture of most Exynos mobile SoC series. His recent interesting point is memory and storage system architecture for AI and Big data system with various new media memory.