18 – Multi-Core, Main-Memory Joins: Sort vs. Hash Revisited

1. Overview of Idea
"Sort vs. Hash" has been famous topic of join operator for a long time. Since hardware and optimization technique have been improved steadily, the results of research also have been changed. This paper includes state-of-the-arts sort-merge join algorithm and comparison of various sorting algorithm such as mpsm, m-pass, m-way. Also, it has comparison of radix hash join and m-way sort-merge join.

2. Main Finding

   Although SIMD instruction and hardware have been improved, hash join is still superior than sort-merge join. Its performance gap can be reduced by increasing input relation size massively.

3. Systems used and its Specifications
Three sort algorithms(m-way, m-pass, mpsm) are used to compare the performance. Then, hash join(radix hash join) and sort-merge join(m-way join) are used to compare the performance of two different join schemes. Base system is columnar model.

4. Workloads evaluated
custom workloads are used, varied on relation size, number of threads, key distribution…….