09 – Integrating Compression and Execution

In Column-Oriented Database Systems

1. Overview of Idea
   data compression on column-oriented database has great advantages because of consecutive similarity of its data. Since there are multiple compression schemes, each schemes has different trade-off, so we should take different schemes at proper situation. Also, light weight operator that can execute operation on compressed data can improve DBMS's performance

2. Main Finding

   Data compression technique can improve column oriented database's performance.

3. Systems used and its Specifications
   Columnar storage structure, using different data compression schemes
   A. Null suppression : compress consecutive null data
   B. Dictionary Encoding : map repeated pattern to some short value
   C. Run-Length Encoding : repeated value can be compressed to (value, length, starting point)
   D. Bit Vector Encoding : maintain bit vector to indicate values' position
   E. Lemper-Ziv Encoding

4. Workloads evaluated
   Custom benchmark and TPC-H are used.