Wyatt Blair

EMT-678-WS: Big Data Technologies

10/11/2024

**Class 4 - Column Store vs. Row Store Abadi et al**

This paper examines the differences in performance between Column-Store DBMSs and Row-Store DBMSs. To demonstrate the magnitude of difference in performance, the authors opt to simulate a column-store DBMS in a row-store DBMS. Then, through use of a benchmark called STAR, the authors examine the difference in performance between the two database styles in apples-to-apples comparison. This benchmark is essentially a standardized schema with a scalable set of data associated with it. Then by testing several different operations on the data across this schema, researchers can gain a benchmark of how quickly their DBMS strategy stacks up against others.

The authors spend some time discussing the differences in optimization strategies available to column-store DBMSs which are not available to row-store DBMSs. Among these strategies are: late materialization, block iteration, column-specific compression techniques, and, (a strategy proposed in this paper) invisible joins.

In their conclusion the authors point out that the takeaway from their work is not that simulating a column-store in a row-store is impossible. Rather that the modern row-store DBMSs poorly simulate column-store systems.