

DATA WAREHOUSE or INFORMATION BLACK HOLE?

Index Inertia Impedes Information Access

For many months and in many articles, the Gartner Group has been trying to alert the computer industry to some potential risks with On-Line Analytic Processing (OLAP) and Data Warehouse offerings that could eventually hinder information access instead of facilitating it.

Their concern is that “information access” vendors are still relying on costly data restructuring technologies that necessitate the ‘indexing’ of individual records in order to provide access to relationships woven among many millions of records.

Though indexing schemes may be a valid approach while databases are small and localized, they are not ‘scalable’ and do not support the ‘interoperability’ required for accessing masses of distributed data.

The only viable solution is one that is scalable (does not suffer as the volume of data swells) and one that supports interoperability (dynamic access to heterogeneous data and distributed platforms). Indexing schemes can not support either of these needs.

One solution is *Set Processing*, a “Data Transformation Algebra” that works on all existing data structures. Set Processing operations access many-records-at-a-time instead of just a-record-at-a-time. These operations also provide more information access functionality than can any data structuring approach.

Unlike index accessing schemes (born in the 1960s and not designed to take advantage of 100 megabyte per second access speeds), Set Processing is specifically intended to optimize I/O performance. Set Processing is intrinsically faster than ‘indexed access’ for mass data access. (Benchmarks against commercially available products have shown that Set Processing typically provides at least an order of magnitude improvement in overall performance.) Set Processing is ‘scalable’ and requires no restructuring of any existing data, (including flat files or legacy VSAM files).

The Set Processing alternative to indexing schemes supports at least one solution to the Gartner Group’s concern. For, unlike indexing schemes, the Set Processing approach performs linearly as data increases and supports parallel information access over multiple heterogeneous platforms (even between disparate Data Warehouses).

If the concern of the Gartner Group is indeed legitimate and goes unheeded, the ‘inertia’ inherent in ‘indexing’ masses of data might eventually preclude timely access to all information.

Today’s Data Warehouses could become tomorrow’s Information Black Holes.