

PASCAL

A Call for Knowledge and Reason

I don't like the dbdebunk.com stuff either. They spend too much time talking about how ignorant everyone is and too little time talking about real practical solutions.

(Horster, Slashdot.org)

Ponder this for a moment, if you will: if most *are*, indeed, ignorant, how can they recognize and appreciate the right practical solution, if and when provided? Indeed, the industry is chockfull of incorrect solutions which are accepted as appropriate, without any questioning, while truly better solutions are ignored, or dismissed as "just theory" because the education and capacity to comprehend them are lacking. In fact, the general practical solution to data management was provided in 1969, but to this day they don't get it. The readers of my site may be aware of my previous Comments On an Interview with Jim Gray (www.dbdebunk.com/page/page/638948.htm), where I gave Gray some benefit of the doubt, on the possibility that he was misinterpreted by the interviewer. I was just made aware by several readers of an article called *Call to Arms*, bylined by Gray with one Mark Compton. Now, I did not expect anything different or better than previous pronouncements, and the fact that Compton is a *marketing* consultant (who calls his company Hired Gun Consulting), does not exactly inspire confidence. My instinct proved accurate: I stopped reading after only a few paragraphs. I don't know who actually wrote the article, but if Gray lends his name to this kind of stuff, he ought to be held to account.

There is so much nonsense in those few initial paragraphs that I am not going to debunk all of it, which would require more time and space that it merits (for an example see *Slashing a Slashdot Exchange* (www.dbazine.com/ofinterest/oi-articles/pascal23, Parts 1-4). Life is too short. Let me illustrate with just a few examples.

We live in a time of extreme change, much of it precipitated by an avalanche of information that otherwise threatens to swallow us whole. Under the mounting onslaught, our traditional relational database constructs—always cumbersome at best—are now clearly at risk of collapsing altogether.

Well, most of the change is either just for the sake of change, no change at all, or even regression: the industry prefers to migrate from ludicrous fad to ludicrous fad, instead of addressing fundamental issues. Every few years it comes up

with some new "paradigm shift" that hacks around problems, unaware that it is often nothing but an old discarded fad that failed.

"Traditional relational database" constructs? Exactly *what* relational constructs, pray? The relational model is logic and math applied to database management. Does Gray mean that *logic and math* are collapsing? Does he mean and database management has a foundation better than logic and math? If so, what is it?

As a participant in the System R project at IBM, doesn't Gray know that the tradition is *SQL, not relational*? And that, therefore, the cumbersome constructs are the result of the SQL team *flouting, ignoring, and violating* relational constructs and good language design? (see *HAVING a Blunderful Time*, or *Wish You Were WHERE?*) Shouldn't he know the difference? And if he does not, what confidence should we have in his pronouncements, let alone his current efforts at Microsoft?

In fact, rarely do you find a DBMS anymore that doesn't make provisions for online analytic processing. Decision trees, Bayes nets, clustering, and time-series analysis have also become part of the standard package, with allowances for additional algorithms yet to come. Also, text, temporal, and spatial data access methods have been added—along with associated probabilistic logic, since a growing number of applications call for approximated results. Column stores, which store data column-wise rather than record-wise, have enjoyed a rebirth, mostly to accommodate sparse tables, as well as to optimize bandwidth.

Here the authors "do a Celko": they throw indiscriminately all sorts of concepts at the reader that have little bearing on the issue at hand, but are hopefully impressive enough for an uninformed audience to accept without questioning. Clustering (physical) are lumped together with time series analysis and decision trees (logical). Besides, are all these new? What is the evidence that the relational model cannot accommodate them? Has he bothered to read, for example, *TEMPORAL DATA AND THE RELATIONAL MODEL*? Do cases that require probabilistic logic imply that we need to discard non-probabilistic logic altogether? Or that we need to dismiss logical inferencing just because we also need to search text?

That vendors add all sorts of features to DBMS products is not proof that they are needed or correct solutions (more likely, just the opposite) to real needs, or even that they are *database* issues. But let's assume for the sake of argument that they are.

That SQL or other non-relational products do not handle them (and maybe are even incapable of it), is that a *relational* fault? What is the *evidence* that XML fares any better?

Let me get it straight: Chamberlin, Gray & Co. failed to implement the relational model, preempting the ability of SQL and its direct-image implementations to provide the very capabilities they find desirable, among them extendibility and flexibility, and now they are blaming *the model* for their failure? What is more, are they seriously proposing the old hierarchic solution that the model made obsolete, and which is even less sound than SQL? Please, give me a break.

Both the earlier interview and the current article constantly and thoroughly confuse the logical and physical levels. Relational technology, has absolutely nothing to do with "access methods", clustering, and other such physical implementation details, and intentionally so. That, and the declarative nature of a truly relational data language make true RDBMSs the *most* optimizable (due to what Codd called "analyzability of intent"), and amenable to multi-processor platforms.

How data is physically stored-column-wise, row-wise, or any other way-has nothing to do with relational constructs. Has Gray heard of *data independence*? Indeed, it is a major point of the relational model to give the implementer complete freedom of storage and access methods (and of changing them, when necessary), without impacting users and applications. This is precisely the difference between true innovations such as the *TransRelational Model* of implementation, and traditional direct-image SQL implementations, which necessitated considerable redundancy (indexes) to achieve good performance, and complicated database administration. So when Gray asks, "Is it any wonder classic relational database architectures are slowly sagging to their knees?", it is impossible to take him seriously.

But wait ... there's more! A growing number of application developers believe XML and XQuery should be treated as our primary data structure and access pattern, respectively. At minimum, database systems will need to accommodate that perspective. Also, as external data increasingly arrives as streams to be compared with historical data, stream-processing operators are of necessity being added. Publish/subscribe systems contribute further to the challenge by inverting the traditional data/query ratios, requiring that incoming data be compared against millions of queries instead of queries being used to search through millions of records. Meanwhile, disk and memory capacities are growing significantly faster than corresponding capabilities for reducing latency and ensuring ample bandwidth. Accordingly, the modern database system increasingly depends on massive main memory and sequential disk access.

Beliefs are better left to religion (which seems to take over American life these days and, worse, displace science).

Application developers "believe" all sorts of things because they lack proper foundation knowledge. It is the responsibility of scientists – be they academics, or in commercial research – to educate them, and to provide sound solutions that dispel unfounded beliefs, not to reinforce or exploit their ignorance. For if the goal is just to give application developers their laundry list of features, we do not need research and scientists. If vendors gave developers science-based TRDBMSs rather than SQL and MySQL, we would not be talking now about "relational deficiencies" and XML. Hardware performance has nothing to do with this (the level confusion raises its ugly head again). Besides, in the quest for performance, should we ignore soundness? (see *The Costly Illusion: Normalization, Integrity and Performance*, www.dbdebunk.com/page/page/1103793.htm).

I am often accused that I mostly "make fun of newbies". Aside from the fact that when, on occasion, I do that, the point is to draw attention to how novices enter the data management field – without proper education, and unwilling to undertake any – the accusation is false. Most of my criticism is targeted at vendor personnel, consultants, the trade media, and academics who are considered, or consider themselves expert enough to issue public pronouncements. If that's what experts like Gray say, what can we expect, under the circumstances, from the average practitioner, let alone the beginner? I wonder what would Ted Codd, a recipient of the Turing award for the relational model, think of Gray's receipt of the same award, given such pronouncements. The industry is so dumbed down, that it can no longer distinguish between novel applications of knowledge and reason, and sheer hacking or marketing. Engineering is not the same as science, and the two should not be confused.

Had knowledge and reason been employed by the US government, we would not be in the mess we find ourselves today. The reader should ponder, in this context, the significance of an American "distinguished" engineer joining a marketing "hired gun" to issue a "call to arms", while ignoring knowledge and reason.

Fabian Pascal

Fabian Pascal is onafhankelijk IT-analist, consultant en auteur gespecialiseerd in data management. Zie ook zijn website www.dbdebunk.com

Special Offer

Author Fabian Pascal is offering DB/M readers discounted subscriptions to the PRACTICAL DATABASE FOUNDATIONS series of papers.

To take advantage of this offer, contact him via the About page at www.dbdebunk.com.