TDWI HOT TOPIC

INVESTIGATIVE COMPUTING: THE NEW FRONTIER IN BI

// BY STEPHEN SWOYER

Today, the practice of "embedding" BI into operational processes yields results that are less analytic than reportorial. The New Model—"investigative computing"—will fundamentally change this.

The way business people expect to consume and analyze information is changing, and business intelligence (BI) is doing its best to keep pace. Perhaps not as rapidly as some line-of-business stakeholders might like, but still pretty quickly, especially by historical standards.

More important, argues industry luminary Colin White, president and founder of information management consultancy BI Research, the very nature of data analysis and exploration is changing.

"If you look at traditional BI and data warehousing, the life cycle is [that] we extract information from operational systems into a data warehouse and only then do we try to run analytics on it," White observes. The New Model is something he calls "investigative" computing. "What we're moving toward is more of this analyze-[and]-store model whereby we blend in all of the data [from a diversity of sources] into an 'investigative' or 'discovery' platform."

White sees this as a complement to the traditional enterprise data warehouse (EDW): although the EDW doesn't go away, it likewise isn't as central to business discovery as it used to be. In the investigative computing paradigm, the EDW becomes a repository for the persistence of analytic insights and a mission-critical system for feeding these insights in real time to thousands of users and applications, in addition to other (more traditional) kinds of data. It's no longer the only central repository for BI and analytics; instead, its role changes to be more "operational" in nature. Information is ingested and processed in other platforms (staging areas) before it gets fed into the data warehouse.

Investigative computing emphasizes a test-driven approach. It's iterative: with each and every new iteration, models get tweaked and algorithms get perfected. High-performance technologies such as in-memory massively parallel processing (MPP) databases can significantly reduce the intervals between iterations, making it easier to train and perfect models and algorithms. So, too, can Hadoop, White concedes—although he stresses that Hadoop makes sense for certain very specific kinds of workloads. These technologies likewise make it possible to analyze ever-larger data sets.

This ability to quickly test models and algorithms against very large data sets—thus keeping iterative intervals short—is a Very Big Deal, according to White.

"As you get these more sophisticated capabilities, you can improve the algorithms. Because we have improved speed technology-wise, we can also run the models faster. We can run a risk model perhaps in a matter of minutes or hours, whereas before it might have taken days or even weeks.

"Based on this [ability to run the models faster], we have the capability to improve our existing models. Second, we may also discover new models or alternative analytic approaches to analyzing data which lead to new insights, or [ways to incorporate] new types of data [into models]. We can then use this to improve the existing production data warehouse."

The key is to identify business needs or drivers before selecting a data- or query-processing power plant, he continues. "If what you want to do is to run analytical models against huge amounts of data and scan it, then Hadoop lends itself to that. You have to look at it workload by workload. On the non-relational side, we tend to identify just Hadoop and [the Hadoop Distributed File System]. There are other things out there, however. We have graph databases, for example. With [investigative computing], you're approaching business problems on a workload-by-workload basis."

In investigative computing, it's no longer appropriate to speak of a data warehouse-driven BI and analytics model. BI in a

sense gets empowered: today, for the most part, the practice of "embedding" BI into operational processes yields results that are less analytic than reportorial.

Investigative computing will fundamentally change this, White predicts.

"As we improve [the power and accuracy of] our analytics, we can actually start to embed them in operational processes. The concept then becomes that we can 'embed' BI ... in operational processes," he explains. "The results of that [investigation] can go in different places. We may put it into the enterprise data warehouse—but we don't necessarily have to."

White's view isn't an outlier. At TDWI's recent World Conference in Las Vegas, Bob Eve, executive vice president of marketing with data virtualization specialist Composite Software Inc., described what he called a "scientific-method-like" approach to analysis. "It's interesting to see the evolution from that more 'deterministic' business intelligence approach—for example, 'I need a report that's like X'—to 'I have this idea I was wondering about, or this hypothesis that I'd like to test.' It's more of a scientific-method-like approach to analysis." Eve, too, sees this as a Very Big Deal.

"I think there's this link between the scientific method and discovery. For too long, [BI] hasn't done anything about addressing this [link]," he noted.

Stephen Swoyer is a contributing editor for TDWI.

ABOUT OUR SPONSOR

TERADATA. THE BEST DECISION POSSIBLE™

teradata.com

Teradata is the world's largest company focused on integrated data warehousing, big data analytics, and business applications. Our powerful solutions portfolio and database are the foundation on which we've built our leadership position in business intelligence and are designed to address any business or technology need for companies of all sizes.

Only Teradata gives you the ability to integrate your organization's data, optimize your business processes, and accelerate new insights like never before. The power unleashed from your data brings confidence to your organization and inspires leaders to think boldly and act decisively for the best decisions possible. Learn more at teradata.com.

ABOUT TDWI

tdwi
THE DATA WAREHOUSING INSTITUTE

tdwi.org

TDWI, a division of 1105 Media, Inc., is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, on-site courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive website, tdwi.org.