Visualizing Big Data Beyond the Warehouse

In an age of Big Data, the adage “Every company is a data company” is truer than ever. The problem is most companies equate the term only with structured repositories and dashboard-style reports; they don’t realize how much valuable data they’re actually sitting on, much less how to rapidly explore and understand all of this information.

Until recently, data sources were a limited universe: databases, documents and spreadsheets, data feeds such as EDI streams and industry-dependent specialized sources like machine-generated information from factory equipment or medical devices. But most organizations could also glean valuable insight from a variety of other sources including PDFs, content management systems, and data feeds for managing equipment and log files.

Analyzing dynamic data presents tougher challenges since the simplistic tools typically available in spreadsheets and other traditional business intelligence (BI) tools are not designed to handle fast updates to the underlying data.

Today, data flows into organizations at unheard-of speeds, from dynamic, real-time sources including ticker feeds, sensors, message busses and complex event processing engines. This data travels networks that are faster than ever. Getting a real-time perspective of what’s happening throughout the organization requires that this data be ingested and analyzed at the velocity it arrives, before landing in a structured repository.

With more data coming faster, pie chart snapshots don’t cut it. Today’s BI tools can’t easily access streaming data or deal with unstructured information. They may do basic visualizations, but the results are static aggregates, limiting the ability to drill down and uncover important, granular information.

Worse, simplistic visualizations built on aggregated data can be dangerous because they make it easy to “dress up” data to support a particular point of view, and equally easy to develop deceptive, incomplete or slanted analyses. Unless users are careful, they can make potentially disastrous decisions based on poorly presented and understood data.

New software technologies are now emerging that let businesses spend less time structuring data and more time analyzing it. They let companies create a visual paradigm for both intake of data and interactive data discovery. More important, the data can be gathered from across the company, bringing in disparate sources of information for a holistic view; tapping into and visually exploring unstructured, semi-structured and structured data simultaneously, in large volumes and at any speed, including true real time.

The new generation of dynamic data discovery visualization provides users with a “lean forward” experience that encourages interactivity. It can visualize data “in flight,” before it gets warehoused, and let users reorganize data on-the-fly, change hierarchies, filter out irrelevant information, and isolate the surprises they find.

**Visualization for Humans**

The ultimate goal is to enable next-generation analytics for fast decision making in a visual data discovery environment designed to work with the way the human brain processes information. Human factors are, in fact, the most important things to be taken into consideration when designing a visualization system.

One interesting area being incorporated into new visualization interfaces is pre-attentive processing; the visual properties that people process almost subconsciously without the need for focused attention. The human mind has an ability to intuitively and effortlessly always know a big shape from a smaller, understand that two shades of red are different and feel that items displayed grouped in a box are of the same kind. Using this power, a load is taken off the short-term memory which otherwise would have a lot more to keep track of through conscious, intellectually costly processes.

By assigning appropriate information value to visual characteristics, including size and color, our perception and understanding of data is greatly enhanced. For example, the automated and effortless perception of information is what visualization tools, such as treemaps, make use of.

The promise of “Big Data” has driven organizations to rethink their approach to traditional business intelligence. Relying on canned reports and static dashboards derived from a relatively small subset of business critical information is no longer an option. To stay competitive, organizations need to harness all relevant information regardless of its type, volume or speed of delivery. Combining real-time data “in flight” with, unstructured and semi-structured data in interfaces optimized for pre-attentive processing and similar human cognitive abilities, next-generation analytics will dramatically improve how Big Data gets visualized to isolate and resolve problems as they occur, perceive hidden patterns, track emerging market trends, and identify opportunities for competitive advantage and improved business processes.

**DataWatch**

www.datawatch.com