Viewpoint

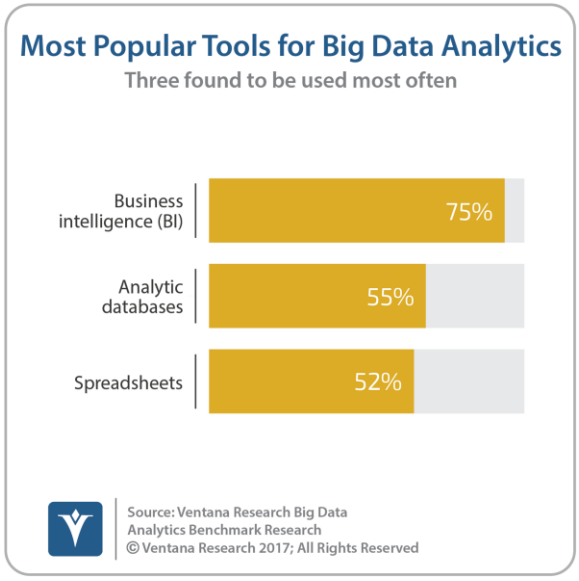
David Menninger
SVP &
Research Director

Empower Business Intelligence for Big Data

Organizations of all sizes and in all industries are being deluged by a torrent of data coming from a multitude of sources in many formats. Typically this data is stored in an array of disparate systems. A key technology for processing and analyzing it all is called big data. If your organization is not exploring big data projects, it should be. Four out of five organizations participating in Ventana Research's benchmark research said that analytics applied to big data is important to their organization. The research also shows that organizations using big data often communicate better, gain competitive advantage, improve efficiencies and respond faster to opportunities and threats.

However, big data also can pose big challenges for users. For example, a primary technology associated with big data implementations is Hadoop, but only one in six (16%) organizations said they have the Hadoop skills available to perform big data analytics.

One path to successfully using big data for business involves giving users access to it through familiar business intelligence (BI) tools. Our research finds that three-quarters of business people dealing with big data use BI tools while half (52%) use spreadsheets. However, it's not enough to simply put a BI tool in front of big data. Processing such volumes of data overwhelms many BI tools, leaving users disappointed by the tool's performance or confused by the lack of organization of the data.



Disappointment also can be rooted in a lack of either the skills to use a specialized tool or an inclination to learn them. Not surprisingly, business users want to work in an environment with which they are comfortable. These users often think in terms of rows and columns, but they want to be able to transpose those rows and columns at any point in their analysis, hence the popularity of spreadsheets and pivot tables within them. Users also approach data with the expectation that it has natural hierarchies: For instance, stores are in cities, cities are in states or provinces, states and provinces are in countries, and these various geographies organize into sales regions. Again, pivot tables within spreadsheets allow users to expand and collapse levels of the hierarchies. Looking at data by hierarchical dimensions seems natural to these users – but data seldom gets recorded and organized this way.

This style of hierarchical navigation and interacting with data is often referred to as online analytical processing (OLAP) or multidimensional analysis. The data model associated with

OLAP is called a cube, like a Rubik's Cube, since users can pivot and reorient the display to show side by side columns that are most relevant to their analysis. Of course the data stored in a database or Hadoop looks nothing like a cube. OLAP or multidimensional analyses can be difficult to implement with standard relational queries implemented in SQL because SQL is predominantly oriented toward rows of data, not cubes. The rows of a SQL table are not easily interchangeable with its columns. MDX, a multidimensional language, was developed to address these issues and support richer analysis of data in cubes.

Most business intelligence tools have incorporated OLAP-style display and navigation techniques using SQL and/or MDX; many tools support one or the other, but very few support both. They have also implemented behind-the-scenes performance enhancements to address the fast interactive response times users expect. Often the data is copied in memory into a columnar structure to speed up interactions with the data. These performance enhancements work reasonably well for moderate volumes of data, but this approach breaks down in a big data scenario – there simply is too much data, and it changes too often to copy and fit it into memory. If users are to be able to continue to work with their BI tools as expected on the big data volumes being dealt with today and expected in the future, an alternate approach is required.

Using a semantic layer – a business representation of data – can make big data look like a cube to standard BI tools. It enables business analysts and modelers to describe virtual cubes in terms of measures, dimensions and hierarchies. The hierarchies in the virtual cubes can also provide mapping to aggregate data; when data volumes are significant, aggregating data for queries is critical to speed up response times. A well-designed semantic layer can also keep track of which aggregations must be updated when the underlying data changes; this helps ensure consistent performance of the BI tools. A semantic layer can also provide the mapping between the big data source and a variety of BI tools so they can all operate on the same data.

The upshot is that working with big data need not be a challenging proposition. It may be tempting to solve performance problems by restricting users, sacrificing functionality and making copies of the data to support specific tools. I advise that you set your sights higher. Move forward with your big data analytics by supporting the BI tools and spreadsheets that your users know and prefer. Big data sources, whether on-premises or in the cloud, are too big to copy, so analyze them in place using a semantic layer that makes the big data source look like and act like a cube. Explore new technologies that provide these capabilities. Your users will be happy, and your organization will be able to start reaping the benefits and value that big data has to offer.

David Menninger – SVP and Research Director, Ventana Research
David Menninger is responsible for the overall direction of research on data and analytics technologies at Ventana Research. He covers major areas including artificial learning and machine learning, big data, business intelligence, collaboration, data science and information management along with the additional specific research categories including blockchain, data governance, data lakes, data preparation, embedded analytics, natural language processing (NLP) and IoT.