

## INTERACTIVE SESSION: TECHNOLOGY

### BIG DATA, BIG REWARDS

Today's companies are dealing with an avalanche of data from social media, search, and sensors as well as from traditional sources. In 2012, the amount of digital information generated is expected to reach 988 exabytes, which is the equivalent to a stack of books from the sun to the planet Pluto and back. Making sense of "big data" has become one of the primary challenges for corporations of all shapes and sizes, but it also represents new opportunities. How are companies currently taking advantage of big data opportunities?

The British Library had to adapt to handle big data. Every year visitors to the British Library Web site perform over 6 billion searches, and the library is also responsible for preserving British Web sites that no longer exist but need to be preserved for historical purposes, such as the Web sites for past politicians. Traditional data management methods proved inadequate to archive millions of these Web pages, and legacy analytics tools couldn't extract useful knowledge from such quantities of data. So the British Library partnered with IBM to implement a big data solution to these challenges. IBM BigSheets is an insight engine that helps extract, annotate, and visually analyze vast amounts of unstructured Web data, delivering the results via a Web browser. For example, users can see search results in a pie chart. IBM BigSheets is built atop the Hadoop framework, so it can process large amounts of data quickly and efficiently.

State and federal law enforcement agencies are analyzing big data to discover hidden patterns in criminal activity such as correlations between time, opportunity, and organizations, or non-obvious relationships (see Chapter 4) between individuals and criminal organizations that would be difficult to uncover in smaller data sets. Criminals and criminal organizations are increasingly using the Internet to coordinate and perpetrate their crimes. New tools allow agencies to analyze data from a wide array of sources and apply analytics to predict future crime patterns. This means that law enforcement can become more proactive in its efforts to fight crime and stop it before it occurs.

In New York City, the Real Time Crime Center data warehouse contains millions of data points on city crime and criminals. IBM and the New York City Police Department (NYPD) worked together to create the warehouse, which contains data on over 120 million criminal complaints, 31 million national crime

records, and 33 billion public records. The system's search capabilities allow the NYPD to quickly obtain data from any of these data sources. Information on criminals, such as a suspect's photo with details of past offenses or addresses with maps, can be visualized in seconds on a video wall or instantly relayed to officers at a crime scene.

Other organizations are using the data to go green, or, in the case of Vestas, to go even greener. Headquartered in Denmark, Vestas is the world's largest wind energy company, with over 43,000 wind turbines across 66 countries. Location data are important to Vestas so that it can accurately place its turbines for optimal wind power generation. Areas without enough wind will not generate the necessary power, but areas with too much wind may damage the turbines. Vestas relies on location-based data to determine the best spots to install their turbines.

To gather data on prospective turbine locations, Vestas's wind library combines data from global weather systems along with data from existing turbines. The company's previous wind library provided information in a grid pattern, with each grid measuring 27 x 27 kilometers (17 x 17 miles). Vestas engineers were able to bring the resolution down to about 10 x 10 meters (32 x 32 feet) to establish the exact wind flow pattern at a particular location. To further increase the accuracy of its turbine placement models, Vestas needed to shrink the grid area even more, and this required 10 times as much data as the previous system and a more powerful data management platform.

The company implemented a solution consisting of IBM InfoSphere BigInsights software running on a high-performance IBM System x iDataPlex server. (InfoSphere BigInsights is a set of software tools for big data analysis and visualization, and is powered by Apache Hadoop.) Using these technologies, Vestas increased the size of its wind library and is able to manage and analyze location and weather data with models that are much more powerful and precise.

Vestas's wind library currently stores 2.8 petabytes of data and includes approximately 178 parameters, such as barometric pressure, humidity, wind direction, temperature, wind velocity, and other company historical data. Vestas plans to add global deforestation metrics, satellite images, geospatial data, and data on phases of the moon and tides.