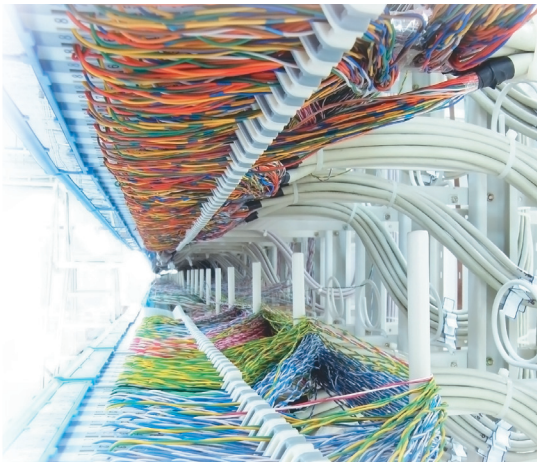

Bringing Big Analytics to the Masses

Neal Leavitt

Big data analytics is valuable to many companies but has been too complex and expensive for smaller businesses. This is beginning to change.

Businesses have greatly benefited from data analytics. Companies analyze activities such as sales, marketing, fraud and risk management, inventory optimization, and customer support to improve their strategic and tactical business decisions.

However, analytics that is powerful enough to work with big data has been too expensive, complex, and compute- and resource-intensive for smaller companies to use.

Now, though, vendors and service providers are leveraging advanced data-processing, network, and cloud technologies to make big analytics suitable for use by such organizations. Some vendors provide analytics software, while others offer services via the cloud.

The stakes are enormous. Market research firm IDC said the big data market is growing about 40 percent annually, with revenue expected to rise from $3.2 billion in 2010 to $17.2 billion in 2015.

Deloitte Touche Tohmatsu, a professional services firm, predicts that about 90 percent of Fortune 500 companies will have a big data initiative in place this year.

A 2012 study by IBM and *The Economist* news magazine found that firms that apply analytics outperform their peers that don't.

As the "Small Companies and big analytics" sidebar explains, small and mid-sized organizations could benefit from analytics as much as large ones, said Robin Bloor, CEO of the Bloor Group, a market research firm. Such companies are struggling with data volumes growing 62 percent annually, according to market research firm Aberdeen Group.

However, these businesses haven't been able to benefit from high-powered analytics and therefore can't make the most out of their information, said Todd Papaioannou, founder and CEO of Continuuity, a big data application vendor.

"The major inhibitors for smaller companies are cost and a lack of skilled personnel," said Aberdeen research analyst Nathaniel Rowe.

Most of these firms don't have the necessary server, software, and network infrastructures to sufficiently process big data, added Chris Piedmonte, CEO of analytics start-up Suvola.

Now, though, advances in developing applications that work with mobile and other smaller systems have made it possible for many organizations to use big analytics software for the first time.

"Cloud-based analytics solutions provide a more accessible price point for adoption," noted Ed Abrams, vice president of IBM Midmarket Business.

BIG-TIME ANALYTICS

Big data products typically require a large IT staff to administer. They also use multiple expensive servers and include software that can be complex to set up and maintain. Big data software tends to use specialized programming languages such as Apache HiveQL, R, and Python.

Another issue is that big analytics is most valuable when used to study many different information types and sources, said Aberdeen's Rowe.

Because organizations need different systems for structured and unstructured data to get the best results, they would have to buy multiple products, noted Leena Joshi, senior director

Published by the IEEE Computer Society
0018-9162/13/$31.00 © 2013 IEEE

of solutions marketing for big analytics vendor Splunk.

Added Robert Ducher, vice president of marketing for analytics vendor InsightsOne, smaller companies have trouble finding employees to work with big analytics, which uses complex and advanced technology, is a relatively new field, and is changing rapidly. People with the necessary skill sets are typically scarce and demand high salaries,

All these factors have made big analytics products suitable only for large organizations.

A common big analytics approach is Apache Hadoop, an open source framework based on Google's MapReduce programming model for breaking large, complex problems into small chunks. Hadoop uses the processing power of clusters of ordinary servers to tackle tasks too large for any of the individual machines.

However, the extensive Hadoop storage and analytics infrastructure is costly to acquire and set up, and it requires specialized talent and ongoing support and maintenance.

"Adding data integration, business intelligence, and developer tools results in a large and complex software stack," said Omer Trajman, vice president of field operations for big data-applications vendor WibiData.

Another barrier for smaller organizations is that the Hadoop APIs are difficult for most programmers to use and there aren't good tools for deploying, running, and monitoring applications, added Continuuity's Papaioannou.

Large companies sometimes build complex in-house machine-learning tools to conduct analytics, said Jos Verwoerd, cofounder and vice president of business development for machine-learning vendor BigML.

This is prohibitively expensive for smaller organizations, noted Indiana University assistant professor Chung-chieh Shan.

ANALYTICS FOR ALL

Handling big data, noted InsightsOne's Ducher, requires a lot of storage and good networks. And analytical techniques such as machine learning use considerable compute power.

However, the cloud has helped make big analytics possible for smaller businesses by letting service providers handle the analytics.

Customers could use today's inexpensive storage to collect large amounts of data, which they could then send to providers via widely available high-bandwidth networks.

And smaller organizations that want to run analytics software themselves can do so on more powerful servers than in the past because of technical advances in lower-cost multicore processors.

However, cloud-based services have become the more popular route because they're scalable and relatively inexpensive, and they don't require a company to have employees with data-analytics-related skills.

"Cloud-based analytics come ready out of the box, with connectivity to existing systems and standard algorithms and reports available," said WibiData's Trajman.

Software-based approach

In the past, only larger organizations had the money to afford big analytics software, as well as the server infrastructure and skilled personnel to run it.

Now, companies such as Suvola offer integrated systems using hardware and software from selected vendors. This lets smaller organizations buy simpler, more affordable all-in-one systems for which the seller provides maintenance and support.

Also, the desire by many companies to use their smartphones and tablets to run powerful business software is inspiring developers to make big analytics and other applications that can run on such devices, noted Mark Levitt, director of business cloud research with market research firm Strategy Analytics.

Unlike desktop software designed to fill large screens with lots of menu options for experienced power users, developers are designing mobile applications to present the most relevant subset of available information and actions based on the user's needs.

This has yielded analytics software that can run on smaller server systems, as well as on wireless devices.

Vendors offering analytics applications for smaller organizations include big vendors

COMPANIES AND BIG ANALYTICS

There are many ways that small companies could use big analytics, even though they don't have nearly the scope of operations found in bigger organizations.

Currently, smaller companies are using big analytics to target customers by capturing useful information about them, said Martin Hack, CEO and cofounder of Skytree, a machine-learning and predictive-analytics software provider.

Marketing managers for smaller stores who need to increase sales could use analytics to identify more effective product promotions and create offers targeted to specific customers.

Product managers and others concerned with customer retention could utilize analytics to predict the risk of each customer going to other companies and then identify actions likely to keep them loyal.

Small companies could analyze IT expenses to ensure they're not wasting resources because of unused capacity or unnecessary duplication, noted Mark Levitt, senior analyst with market research firm Strategy Analytics.

such as IBM, Oracle, SAP, and SAS, as well start-ups like QlickTech, Tableau Software, and Tidemark.

Continuuity offers a suite of products that let Java developers more easily build, test, and deploy big analytics programs. It also provides the Continuuity AppFabric runtime hosting platform, which lets developers deploy, monitor, and manage the applications.

Cloud-based approach

Cloud-based analytics either accepts uploaded data from a company or connects directly to its website or back-end service. For example, analytics from business-intelligence provider GoodData communicates directly with a company's Zendesk customer-support software by using the API and credentials for Zendesk's back-end services.

Smaller companies haven't used the cloud-based approach for big analytics until now in part because the cloud-based services marketplace is relatively new, noted BigML's Verwoerd. And, he added, "Prices are just now dropping to levels affordable for [them]."

It hasn't always been easy either for smaller companies to transmit large amounts of data over networks or for service providers to return analytics quickly. Technical advances, such as faster networks and processors, have helped with this.

Today's cloud-based big analytics providers include Amazon, BigML, Cloudability, CloudVertical, Cloudyn, Continuuity, BigML, InsightsOne, Newvem, Rackspace, RightScale, and Uptime Software.

Splunk Storm is a cloud version of Splunk's big data analytics software, made possible by the emergence of dependable, cost-effective commercial cloud-computing services, noted Leena Joshi, the company's senior director for solutions marketing. Because Storm is cloud-based, she said, it's easily scalable.

DRAWBACKS

Big analytics for small organizations faces several potential hurdles.

According to the results of an Aberdeen survey in which companies could provide up to three reasons why they haven't invested in big-data-related technologies, 51 percent said the software and services are too expensive, 40 percent said they lack the necessary IT resources, and 37 percent said they don't see a business need for it.

Cloud-based analytics work in silos, which enables them to be tightly designed for one domain.

Cloud-based approaches also suffer from the same challenges that other cloud technologies face, such as cost and scalability.

In addition, noted Suvola's Piedmonte, there is a limit as to how much data cloud-based analytics can handle. "When you're talking about terabytes of data," he explained, "the bandwidth just isn't there for that now. If you are generating a lot of on-site data, moving the data to a remote cloud for analytics isn't going to fly if you need to extract tactical business knowledge in real time."

Security is another issue, as many organizations are hesitant to push their core data into the public cloud, said Continuuity's Papaioannou. Some organizations, especially those in the financial and healthcare sectors, have policies requiring that they store information only on internal systems.

Cloud-based analytics work in silos, which enables them to be tightly designed for one domain, according to WibiData's Trajman. However, he noted, this means they don't typically share data or results across department or disciplines within a company.

When organizations want to do this, he said, they typically must build their own cross-domain analytics system. External cloud-based services typically won't be able to integrate closely enough in an organization's IT architecture, noted University of California, San Diego professor Charles Elkan.

Over time and by reviewing customer input, vendors will figure out ways to make big analytics less complex and thus easier to use.

For example, said Aberdeen's Rowe, analytics software eventually will be able to process requests through a GUI or simple SQL commands rather than via approaches requiring specialized programming or hand coding in new and unfamiliar languages.

Aberdeen's research indicates that self-service integration tools will allow for fast connections to multiple data sources, while interactive reports and dashboards will enable easy visualization and exploration of large datasets.

Products could even emerge that integrate data collection and analysis into daily business activities.

According to WibiData's Trajman, big analytics tools eventually will be able to integrate multiple datasets from across departments and domains within a company to enable deeper analytical insights.

Many of big analytics' future benefits will arise from combining information from multiple sources,

both within an organization and across collaborating groups, added the University of California, San Diego's Elkan.

Of the two approaches, said InsightsOne's Ducher, cloud-based analytics will emerge as the default for small organizations or for departments within big companies.

One potential issue, noted Rowe, is that as even smaller organizations work with increasing data volumes, there may not be the available network bandwidth for quickly and frequently sending information via the cloud to an analytics provider.

Many users who aren't technically savvy might not fully understand analytics' benefits or how to utilize the approach and thus may choose not to employ it.

Nonetheless, said Flavio Villanustre, LexisNexis Risk Solutions' vice president for information security, "Organizations leveraging analytics will have a greater competitive advantage. Those that don't will lag behind their peers."

Neal Leavitt *is president of Leavitt Communications (www.leavcom.com), a Fallbrook, California-based international marketing communications company with affiliate offices in Brazil, France, Germany, Hong Kong, India, and the UK. He writes frequently on technology topics and can be reached at neal@leavcom.com.*

Editor: Lee Garber, *Computer*; l.garber@computer.org

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.

IEEE VR 2013

20th IEEE Virtual Reality

16-20 March 2013
Orlando, FL, USA

IEEE VR 2013 is the premier international conference and exhibition on virtual reality. Find the brightest minds, the most innovative research, the leading companies and the most stimulating discussions in the fields of virtual environments, augmented reality, haptics, and 3D user interfaces.

Register today!

http://conferences.computer.org/vr/2013

IEEE

IEEE computer society