

A Revolution in Analytical Technology

Thomas H. Davenport

It's been ten years since Jeanne Harris and I published the book *Competing on Analytics*, and we've just finished updating it for early fall publication. We realized during this process that there have been a lot of changes in the world of analytics, although of course some things have remained the same. The timeless issues of analytical leadership, change management, and culture haven't evolved much in ten years, and in many cases those remain the toughest problems to address.

But analytical data, technology, and the people who use them have changed a lot. I must confess that I didn't anticipate how difficult updating the book would be (so please buy it when it comes out to make the effort worthwhile!). I thought that perhaps a few "global replace" commands in Microsoft Word would do the trick—change "terabytes" to "petabytes," for example, and "quantitative analysts" to "data scientists." Yes, we did make a few such easy changes. But there are a lot more that required more than simple word substitution.

I won't go into all of them here, but below is an annotated list of changes over the past decade in the world of analytical technology:

- *Big data*—interest in this term started to rise in about 2011, according to Google Trends. Of course it began to take off earlier in Silicon Valley, with the rise of Internet behavior (clickstream) data. These new data sources led to a variety of new hardware offerings involving distributed computing. And the need to store and process this data in new ways led to a whole new raft of software, such as:
- *Hadoop and the open source revolution*—Hadoop was necessary to store and do basic processing on big data, along with such scripting languages as Pig, Hive, and Python. Since then we've seen other open source tools rise in popularity, such as Spark for streaming data and R for statistics. Acquiring and using open source data is a pretty big change in itself, but each of these specific software offerings brought a set of new capabilities.
- *Data lakes*—Data lakes are Hadoop-based repositories of data. These don't perform analytics themselves, but they are a great way to store different types of data—big and small, structured and unstructured—until it needs to be analyzed.
- *Operational analytics*—Many organizations want and need to integrate analytics with their production systems—for evaluating customers, suppliers, and partners in real time, and for making real-time offers to customers. This requires a good deal of work to integrate analytics into databases and legacy systems.
- *Componentization and micro-services*—integration with production systems is much easier when the analytics are performed by small, component-based applications or APIs. Even proprietary vendors like SAS are moving in this direction.

- *Streaming analytics*—Internet of Things and other streaming data sources have made it increasingly desirable to analyze data as it streams into an organization. This often requires integration with some sort of event processing technology.
- *Grid/in-memory analytics*—A big change in analytics has resulted in a change in the hardware environment for computing analytics. The outcome is speed—often order-of-magnitude increases in the speed of doing analytical calculations on data. In many organizations, the idea of submitting your analytics job and getting it back hours later is a distant historical memory.
- *Cognitive technology*—I’ve saved some of the most important technologies for last. A key assumption behind analytics in the past is that they are prepared for human decision-makers. But cognitive technologies take the next step and actually make the decision or take the recommended action. They are actually a family of technologies, including machine and deep learning, natural language processing, robotic process automation, and more. Although most of these technologies have some form of analytics at their core, to me they have more potential for changing how we do analytics than any other technology.

No wonder we’re all tired! Just keeping up with these technologies and updating our analytics infrastructures to accommodate them is a full-time job. In addition, users need to be educated on which technologies make the most sense for their business problems.

And the old technologies haven’t gone away. Companies still use basic statistics packages, spreadsheets, data warehouses and marts, and all sorts of small data. In almost every organization, one can make a case that it’s the combination of the new and the old that makes the most sense. In data storage, for example, the structured data that needs lots of security and access control can reside in warehouses, while the unstructured/prestructured data swims in a data lake. And if you want to understand your customers, you certainly need a combination of big and small data—and the combination of methods and tools to analyze them.

This much change in ten years surely constitutes a revolution. And the combination of new and old analytical technologies in most firms requires that we add more resources to manage them. That’s the downside of this revolution, but the upside is that we have a better understanding of our business environment than at any other time in history. Let’s hope we use it to make some great decisions, take informed actions, and introduce great new products and services based on data and analytics.