



Hadoop and Big Data: The Year Past, The Year Ahead

By John Schroeder, CEO and Co-founder, [MapR Technologies](#)
December 11, 2012

<http://tdwi.org/Articles/2012/12/11/Hadoop-Big-Data-2012-2013.aspx?Page=1>

Technology advances widened Hadoop's enterprise suitability. In 2013, Hadoop will firmly establish its dominance in big data analytics.

During 2012, Hadoop gained significant traction in the marketplace as Web 2.0 firms increased operational scale and deployments across traditional market segments moved from early use-case identification to large production deployments. Technology advances made Hadoop suitable for a more broad range of organizations and their use cases. In 2013, Hadoop will firmly establish its dominance in big data analytics with the addition of even more capabilities.

The growing volume, variety, and velocity of data continue to drive the need for capable and cost-effective analytical tools. The market has responded with a wide variety of solutions targeting both general and application-specific needs, and with a choice of open source and proprietary platforms. Here is my take on the key trends the industry witnessed in 2012, along with my predictions for the three biggest trends Hadoop will experience in 2013.

A Look Back at 2012

Last year, we expected that with Hadoop getting easier to use and easier to manage at scale, we would see more applications in use in more organizations. How popular Hadoop would become in the span of just a single year is remarkable.

2012 Trend #1: Revenue-generating use cases trump cost-saving applications

Hadoop has always been a good fit for applications that process massive amounts of data for predictive modeling and other analytics. Increasingly, these applications are being used to generate revenue by more precisely targeting and in some cases adapting products and services.

Examples of this trend can be found everywhere, including online advertising, retail purchase recommendations, couponing, and new offerings. Ancestry.com, for example, is using Hadoop to increase revenue with an interesting AncestryDNA service that leverages its 34 million family trees and 10 billion records. Hadoop is the ideal platform for applications that require interactive and iterative analytics with multiple algorithms.

A major retailer also found that by taking advantage of Hadoop's scalable storage and processing power, more data could be analyzed more quickly than ever before. Analyses that used to take six weeks to process just 10 percent of the company's sales data can now be done daily on 100 percent of the data with Hadoop.

2012 Trend #2: Hadoop pulled away from the other big data analytics alternatives

The ascendancy of Hadoop has been truly remarkable. Hadoop has distanced itself from MongoDB, Cassandra, Couchbase, and the plethora of NoSQL options to become the safe choice. Support for Cassandra has dropped off, with Facebook reducing its investment in the technology and the realization that an eventual consistency model is appropriate for only a limited set of use cases. MongoDB's growth has flattened despite having a friendly programming environment due to lack of scalability. The reasons for Hadoop's growing popularity are understandable. Hadoop provides the most effective and cost-effective way to capture, organize, store, search, share, analyze, and visualize disparate data sources (structured, semi-structured, and unstructured) across a scalable cluster of commodity servers. In stark contrast to the fractured and niche-oriented nature of the alternatives, Hadoop offers what users really want: a uniform approach to and broad set of APIs for big data analytics (including MapReduce, query languages, and database access with easy integration of leading analytic and search platforms) along with an expanding ecosystem delivering the broadest range of technology and services.

2012 Trend #3: Hadoop expertise grew rapidly, but a shortage of talent remained

Ensuring Hadoop's reign in big data analytics is the growing talent of software developers, data scientists, and operations personnel that is growing fast but not yet keeping up with the demand. Techies gaining Hadoop experience will be able to write their own ticket for the next couple of decades, like being an Oracle DBA in the early 90s. The Strata Hadoop World 2012 event, for example, attracted an impressive 3,000 attendees. Further evidence of Hadoop's dominance was revealed in the job market by indeed.com, which noted that every year since 2008, Hadoop has eclipsed other hot applications, including mobile and social media, with a percentage growth that reached an astonishing 400,000 percent in 2012.

A Look Ahead: What to Expect in 2013

In 2013, Hadoop will continue to expand capabilities and will be used in a growing number of applications, which will further establish its market leadership in big data analytics.

2013 Trend #1: SQL-based tools for Hadoop will continue to expand

A substantial number of data analysts are quite adept at using a structured query language and naturally desire to put this skill to good use with big data, so it should come as no surprise, then, that Hadoop's support for SQL is expanding to accommodate this need. Although this is not a major trend given the growth in the talent pool with native Hadoop skills, it is one worth watching.

One example of a SQL-like language for Hadoop is HiveQL, which converts scripts (existing or new) into MapReduce jobs. Other examples include: Drill with its DrQL query language;

Hadapt, a native implementation of SQL; and Impala, a real-time query engine. They all make Hadoop accessible to the large SQL fluent community.

2013 Trend #2: HBase will become a popular platform for BLOB stores

HBase provides a non-relational database atop the Hadoop Distributed File System (HDFS). HBase applications have several advantages in certain distributions, including the creation of a unified platform for tables and files, no need for splits or mergers, centralized configuration and logging, and consistent throughput with a low latency for database operations. Some distributions also add support for high availability, data protection with mirroring and snapshots, automatic data compression, and rolling upgrades.

One application that is particularly well suited for HBase is BLOB stores, which require large databases with rapid retrieval. BLOBS -- binary large objects -- are typically images, audio clips or other multimedia objects, and storing BLOBs in a database enables a variety of innovative applications. One example is a digital wallet where users upload their credit card images, checks, and receipts for online processing, easing banking, purchasing, and lost-wallet recovery.

2013 Trend #3: Hadoop will be used more in real-time and lightweight OLTP applications

With its roots in search engines, Hadoop was purpose-built for cost-effective analysis of datasets as enormous as the World Wide Web. The millions of pages of content are analyzed in batches and then served up during searches in real time. The advances I've mentioned and other improvements in Hadoop's capabilities now make it possible to stream data into the cluster and analyze it in an interactive fashion -- both in real time.

Use cases such as telecommunications billing and, in some cases, logistics applications have outgrown traditional relational database architectures. Telcos 30 years ago tracked a few hundred calls per week to a single home phone. Today they track data and voice transactions to a multitude of devices in one household. Hadoop and HBase provide scale and efficiency advantage for these types of applications over traditional relational data models.

With all of these changes, 2013 will be remembered as the year Hadoop firmly established its dominance as the strategic choice for big data analytics.

About the Author

John Schroeder is CEO and co-founder of [MapR Technologies](#). Schroeder has led companies creating innovative and disruptive business intelligence, database management, storage, and virtualization technologies at early-stage ventures through success as large public companies. He was previously CEO of Calista Technologies (Microsoft), CEO of Rainfinity (EMC), and senior vice president of products and marketing at Brio Technologies. Schroeder has also held general management and executive positions at Compuware, Candle, and SAIC. You can contact the author at John@maprtech.com.