

Avoiding the Magic Unicorn Syndrome with Big Data

Authored by Doug Slemmer, President and CEO, at iOLAP, Inc.

We've all read the articles where someone with a massive data set unleashes the power of Big Data and discovers an awesome and magical insight about their customers that changes the course of history as we know it. We like to call them *Magical Big Data Unicorns*—so frequently discussed and pursued, but so rarely seen in the wild! We're not saying they don't exist, but we would like to suggest a better approach to getting solid ROI from these new-fangled Big Data tools.

For example, when many data gurus think of Hadoop, they think of analyzing huge, unstructured datasets that will uncover important corporate insights. It feels like everyone is marketing the power of Hadoop to find that priceless gold nugget buried within petabytes of social media, blogs, emails, reviews, clicks and chats. Because those stories have great anecdotal power behind them, they can often light the imaginations of business users and help vendors sell new products. Just like a Magical Unicorn.

What is not as frequently advertised and discussed is the ability for Hadoop to help organizations save really big IT dollars in far less dramatic ways. These scenarios don't make for cool infographics or fascinating Nate Silver data epiphanies, but they are absolutely meaningful techniques for significantly reducing IT costs.

This article examines several scenarios for using Hadoop to save on IT budgets and improve your Service Level Agreements (SLAs). While there are other possibilities of cost savings, this article covers the following scenarios where Hadoop-based cost reductions could be implemented and measured.

- Faster transformations that allow “crunch-time” processing to be offloaded to less expensive resources and help reach SLA objectives
- Create “active archives” for valuable but infrequently accessed data at a far lower price point than your expensive database but far more accessible than traditional archival methods

- Create real-world SQL “sandboxes” in Hadoop that are less expensive than current sample-size testing and sandbox environments.

Hadoop will not replace an organization's enterprise data warehouse. However, more and more enterprise data warehouse architects now believe that Hadoop could assume pieces of the data warehouse workload—and with that “joint architecture” come significant cost savings.

Faster (and Cheaper) Transformations

As data volumes grow, and data types increase and become more diverse, SLAs for various projects and phases are put into jeopardy. IT staff are constantly squeezing everywhere they can for more time and cycles to get everything done in increasingly tight 24 hour windows. In fact, huge data users like Yahoo, LinkedIn, Facebook, Twitter and Google, became the key drivers for development of the open source Hadoop community in the first place. They simply needed a wider-scale, faster platform. Early adopters at Yahoo tell of transformations that used to take eight hours shrink to 15 minutes due to Hadoop performance capabilities.

Hadoop has the ability to allow IT staff to offload highly compute-intensive manipulation and transformation tasks at a much lower cost than typical data warehousing platforms. Data transformation is an excellent candidate to use the power of Hadoop's processing to offload data these data manipulation workloads. This frees the data warehouse to more effectively execute queries, perform analytics tasks and generate timely reports.

When a data warehouse is paired with Hadoop, the organization can now avoid the trade-off between valuable transform compute cycles and the value returned from an unencumbered data warehouse.

Active Archives

A key benefit to Hadoop is that companies can take advantage of the decline in storage costs across the board. Hadoop can be a very cost-effective alternative to keep older data out of the data warehouse—but still keep it online. These “active archives” could be a benefit to many organizations that might want or need the ability to query multi-year datasets for patterns and outcomes.

Commercial BI tools are now available to use SQL with Hadoop directly—and these tools will continue to evolve and improve. This will accelerate the utilization and interest in these inexpensive active archives for data mining and other data science activities.

Hadoop Sandboxes

Data scientists and analysts were big users of small data samples to test things like predictive analytics. It was an ideal place to explore data. As systems and needs changed, the size of the sandboxes got bigger so that predictions and other results improved over time.

Now, with the advent of inexpensive, large data stores, sandboxes can be used for real-world testing, analysis and simulation. Using the features of Hadoop, predictive analytics can be run against enormous quantities of data. If your organization has been using expensive platforms for these sandbox systems, huge cost savings can be attained by moving them to Hadoop and its less expensive offline storage.

Data scientists are completely engaged in developing new scenarios that let them now analyze unlimited dataset sizes for increased accuracy in their models.

Summary

Hadoop has the ability to significantly lower costs of data management. It will not replace your data warehousing and BI infrastructure. It can help reduce contention for operational resources and allow cost-effective access to active archive data for longer-term historical analysis. One more benefit is the reduced need to continually add costly new technology to the data warehouse simply to meet service level agreements.

Start small when reviewing where it is believed Hadoop can play a role in reducing costs. Find time and resource pain points and see if a solution can be built from that perspective.

About the Author

Doug Slemmer is the President and CEO of iOLAP, Inc., and is responsible for the corporate direction and service quality of the professional development team at iOLAP. Since co-founding iOLAP, Mr. Slemmer has pioneered the concept of a consulting and implementation group that focuses exclusively on Business Intelligence. With over 18 years of industry experience focused entirely on Business Intelligence and Data Warehousing, he brings real-world knowledge and creative leadership to the solutions iOLAP rolls out for its customers. Prior to starting iOLAP, Mr. Slemmer was the Data Warehouse Practice Manager with Clarity Consulting. He is a graduate of the University of Texas at Austin.

About iOLAP

iOLAP, Inc. (www.iolap.com) is a Dallas-based strategic data consultancy specializing in Data Warehousing and Business Intelligence strategy and solutions. iOLAP is completely focused on the Data Warehousing (DW) and Business Intelligence (BI) markets and brings a client-centric and business-focused perspective to all of its engagements. iOLAP has been in business since 1999 with in-depth expertise across all Data Warehousing and Business Intelligence technology areas, including the newest leading-edge Business Analytics and Big Data architectures. iOLAP serves some of the world's largest companies across all major industries and employs over 75 full-time employees based in the United States and over 25 full-time staff in our offshore development center in Croatia. All sales and support comes from our centralized offices in Texas and California.

For More Information

For more information about iOLAP services, call (214) 618-5000 or email us at info@iolap.com. Please visit us at <http://www.iolap.com>.



Business Intelligence + Big Data