



Avoiding Big Failure with Big Data

By Doug Slemmer, President and CEO of iOLAP, Inc.

Companies are being pressured on multiple fronts to do something amazing with Big Data. But should the brakes be applied, even slightly, to consider some lessons from the past? Historically, large data warehouse projects failed at surprisingly high rates. How do we learn from these collective past mistakes and increase the odds that Big Data efforts will provide Big Returns? While new buzzwords and acronyms continue to appear on the strategic data landscape, Big Data projects share most, if not all, of the same risks traditional data warehouse efforts face.

Avoiding the Field of Dreams

If the approach is to simply start building something with Big Data and believe that deep insights and value will then magically appear, you are probably new to strategic data initiatives. This undisciplined approach is by far the single biggest reason for the majority of the large-scale data warehouse failures. Currently there are powerful forces at work that are generating strong momentum around Big Data--momentum that can hurt your employer, and derail your career, if you're not cautious. Hardware vendors, software vendors, service firms big and small, and even the New York Times and Forbes are all shouting about the wonders of Big Data. Gartner suggests that Big Data will create 1.9 million new U.S. jobs in Big Data by 2015. Are we heading towards a Big Data Bubble?

It's time to take a deep breath, pause a few moments, use some common sense, and develop a game plan. A good approach to take is, not surprisingly, similar to how you should approach any data and analytics project (or any IT project for that matter).

Understand the Potential

A good place to start is to understand how Big Data is any different from what everyone was doing last year with data that was, well, also pretty big. This article is not an analysis of why Big Data is all Big Talk.

On the contrary, technologies which are being developed around and within the Big Data ecosystem hold tremendous opportunity for today's business environment--as long as it's understood why Big Data technology is different and how it can give you a competitive edge. Understanding the strengths for any given tool or technology enables you to begin reviewing strategies to exploit those strengths.

Big Data Advantages

Increased Data Fidelity. Webster defines fidelity as: "accuracy in details; exactness." Put simply, the more data points there are to analyze, the greater the accuracy of the results. Drawing conclusions from only ten retail stores versus ten thousand was often all recent database systems could handle. Those ten stores could easily be generating terabytes of T-log (sales transaction) data every year. Those numbers and required investment was music to the hardware and storage vendors' ears. Big Data technology changes that cost equation so you no longer need to debate with the business analysts about how many months of transactional data to keep online. Big Data storage is extremely cheap--so keep it all.

Increased Correlations. It is estimated that 80% of the data that businesses should be looking at is stored outside the corporation. Ralph Kimball has written extensively on the power of data correlations. Finding meaningful and unexpected data relationships that contribute to understanding consumer behavior or improving business performance is what *Business Intelligence* is all about. The case for important correlations becomes critical in healthcare. Patient data has historically been constrained to a handful of basics: symptoms, diagnosis, procedure, and outcome. But what about the mountains of data to which analysts could connect or correlate? Information like demographic and lifestyle information for the patient cohorts, transaction level vitals collected from the moment of admittance, and likely information continuing to stream from medical devices throughout the patient's stay? In addition, remote sensors and monitoring after the hospital stay are just now starting to emerge and will continue to add volumes of valuable data that can be used to identify post-treatment risk factors and procedure benefits. At the end of the day, every business is trying to connect the dots, so having more dots is inherently better.

Increased Analytics Complexity. Have you or someone on your team ever tried to write a SQL statement that answers a really complex business question? For example: find all of the customers who used their credit card in more than three geographic locations within a 24 hour span who do not normally travel and where those transactions were for business categories not typically related to

travel. Oh, and exclude any online vendor they may have previously done business with in the past two years as those would generate false positive hits. Can you write it? Sure, it's a massive sub-query juggernaut with multiple very large temporary tables that will likely make your DBA scream as you bring the data warehouse to its knees. With new Big Data technologies, that same question can be asked and answered with ease and without taxing the system.

Increased Analytics Speed. This one needs the disclaimer of *"it depends"*. Hadoop clusters glued together from discarded older generation servers can take ten times longer to answer the same questions as a mediocre Oracle box-- but the potential for speed is there. The truth is that there are some things that highly optimized, structured relational databases can simply do better. In other situations, the new technologies like HDFS (Hadoop Distributed File System) and MapReduce will absolutely crush traditional platforms. The old adage of "right tool for the job" applies here.

Increased Data Ingestion. Done properly, Big Data technology can dramatically increase the rate at which your systems can ingest data. Streaming data sources are here today and are increasing exponentially. The bigger question to ask is if there is true business value in near-real time data analysis because those projects are not cheap. If no value exists, it is hard to consider it an advantage. But, if streaming analytics is a short or long-term need, the capability is available now.

The Next Step: Monetization

Before you spend the company's first Big Data dollar, find a list of strong opportunities to monetize that data considering the key strengths described above. Push back on the "build it and hope" strategy, even if business and the C-level executives are all but giving you the green light and budget to do exactly that. Early excitement and efforts will never reach full potential without a clear charter of monetization opportunities that should be driving IT and business intelligence investments. Big Data opportunities can be readily apparent, numerous, plentiful, and ready for the picking, or they can be a bit more difficult and hidden. Sit down with the players from both IT and Business Units, and spend the time and effort to carefully identify the best initial Big Data project candidates. Don't be afraid to bring in some outside consulting help who have gone through the early learning curve with early adopter clients.

Outlined below are three categories of opportunities to explore as a team. The list is by no means complete. It simply helps start a conversation and a whiteboard session or two.

Prediction

Big data provides the opportunity to develop better predictive models for the metrics that matter to business. The increased fidelity and correlations concepts mentioned earlier in this article provide ample raw material for sharp data scientists to do contemplate and research. These opportunities can come from Business Units, IT, or Operations, but again everyone needs to open their minds to what is possible. Read a lot of the new literature out there. People tend to have a hard time resetting what is or isn't possible when it comes to technology and expectations. Be prepared to provide examples that are as close to your business as they can be. It often helps to have both vertical (my business) and horizontal (my group) examples to spark imaginations.

Suggestion: Make a list of the key metrics each group is driven by, and review opportunities to leverage Big Data into predictive models for those measures. For example, if Net Promoter Score (NPS) is a Big Deal, find ways to use Big Data to model and predict that customer sentiment.

Optimization

Big data increases the ability to optimize the systems that create Big Data. This *circle of life approach* can be a great place to start looking for monetization opportunities. For communication service providers CDR (call detail record) and IPDR (internet protocol detail record), the transactional level record of every call, text, email, tweet, stream and post is some of the biggest (and best) data around. How to optimize the systems that generate that data is a rich target environment.

Suggestion: Pay close attention to the systems that exist today and generate Big Data. Frequently, that data is not being stored or leveraged for all it's worth. That's probably because the cost/benefit to storing it was previously too low. Just like energy companies developed more efficient ways to extract gas from shale deposits, Big Data technologies can open up previously untapped opportunities to leverage existing Big Data generating systems.

Commoditization

Gartner predicts that by 2016, 30% of companies will be selling or renting their Big Data. Companies may not understand the value that others may find in their data, discarding it as noise, but every bit of data has potential value.

Suggestion: Review what data is being discarded or possibly not even captured, and assess the opportunities that it presents. Initial push-back on privacy or the proprietary nature of your data is warranted but should not be used to kill-off the option before the monetization value is fully vetted.

The very nature of Big Data technology supports the "Fail Fast" model. After you have identified 2-3 possible monetization scenarios, build out quick proof-of-concepts, and quickly determine if the assumptions are valid before moving forward. Be prepared to move the winners of those experiments quickly to the next level, and continue to look for more monetization candidates.

Summary

Big Data offers tremendous opportunities, but like any new strategy, the right tactics need to be wrapped around the strategy. Put some of your best thoughts and efforts into how to use Big Data to a corporate advantage. Do it right and you will be a Big Hero.

About the Author

Doug Slemmer, President and CEO

iOLAP, Inc.

Mr. Slemmer is responsible for the corporate direction and service quality of the professional development team at iOLAP (www.iOLAP.com). 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, Inc.

iOLAP, Inc. is a Dallas-based strategic data consultancy specializing in Data Warehousing (DW) and Business Intelligence (BI) strategy and solutions. iOLAP has been in business since 1999 with in-depth expertise across all DW and BI technology areas, including the newest leading-edge Business Analytics and Big Data architectures. For more information, visit their website (www.iolap.com) or email info@iolap.com.