The 3 questions to ask yourself about

BIG DATA

Do you have a big data problem?

Companies looking to tackle big data problems are embarking on a journey that is full of hype, buzz, confusion, and misinformation.   But if they focus on asking the right questions to solve the right problems, the journey can be much smoother with better results. Gartner predicts that until 2016, confusion around the term big data will inhibit growth and adoption of these tools. According to their report, "Recent Gartner surveys show that only 30 percent of organizations have invested in big data, of which only a quarter (eight percent of the total) have made it into production."

In the world of big data, there's a sort of enterprise equivalent of "keeping up with the Jones's." While many claim to do "big data", few have truly adopted enterprise wide solutions in the way that Google, Facebook, Twitter, and only a handful of other companies have. A lot of companies think they're being left behind by not getting into big data. The first question they seem to be asking is "how can I get started with big data?" But the real question is "can I derive value from my big data?"

One of the biggest things companies overlook is figuring out if they even have a big data problem at all. What is typically missed within all the marketing & buzz is that big data is a different kind of data that requires a different kind of solution. Note that the original use case for big data was to solve a rather unique problem. It started with Google's whitepapers on map-reduce and Google Filesystem (GFS) to solve the problem of storing the contents of every webpage on the Internet and then figuring

out how they all link to each other (page rank). That's not a problem most companies have.
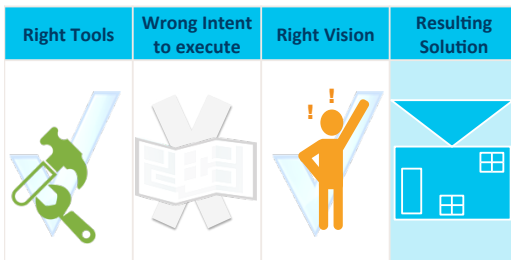
What are your real analytics challenges?

Big data should be considered once the existing analytics solution can't accommodate the new data analysis needs. Companies should ensure that they are using the right approach to develop a big data solution by first understanding the analytical challenges they face. Virtually all companies share a common problem when it comes to building an analytics solution: It's very difficult. Bringing together people, processes, and data from disparate parts of an organization is no simple task. If performing analytics were easy, we could predict the stock market and be rich. If we use building a house as an analogy to building an analytics solution, there are a few equivalent problem areas:

◻ Lack of vision – results in a mess of disintegrated solutions

Right Tools	Right Intent to execute	Wrong Vision	Resulting Solution

- Without a coherent, long term and encompassing vision, the individual components of the house simply don't fit very well together.

◻ Poor execution – results in integrated tools that are unusable

Right Tools	Wrong Intent to execute	Right Vision	Resulting Solution

- Even with the right vision, not executing properly will result in something that may look somewhat like a house but clearly isn't what was intended.

▢ Having the wrong tools – results in components not working together

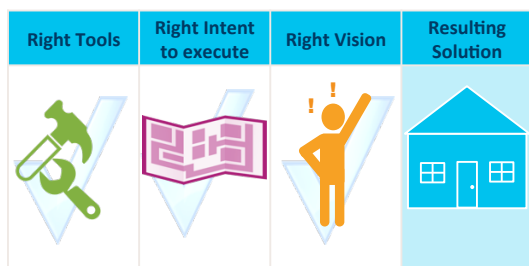| Wrong Tools	Right Intent to execute	Right Vision	Resulting Solution

- While this doesn't tend to happen very often (in fact most organizations have too many tools), having the wrong tools will make it impossible to even start building your house.

When companies run into these problems with their analytical solutions, they're forced to perform manual and time intensive tasks that are prone to human error.
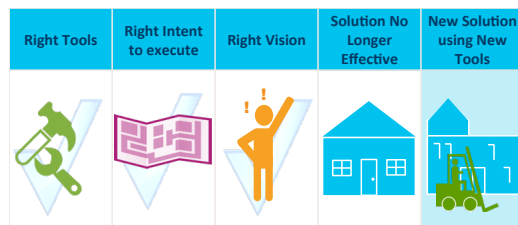
Defining Big Data

As we mentioned, the first thing companies need is to determine before implementing a big data solution is whether or not they have a big data problem. The three V's – Volume, Variety and Velocity are the key to defining and understanding the need for big data.

▢ Volume, or the amount of data addressed by big data, is beyond that of a traditional database. There are no set limits to the volume, but it's generally accepted to be in the Terabytes to Petabyte range. Companies that employ big data solutions will generally look to preserve all data even if there is no immediate use for it. For companies challenged with handling such large volumes, a big data solution will scale horizontally, seamlessly adding more capacity to meet the increasing load.

▢ Variety, or the range and types of data and sources in big data, are endless. Big data solutions process a variety of multi-structured data, such as social media, text, server logs, geo-location, etc. A big data solution can support data storage in "schema-less" architecture to enable a company to analyze a variety of data at extraction, instead of storage. The "schemazation" is performed at extract instead of load.

▢ Velocity is not only the speed with which the data goes in and out, but also the rate with which it changes. Linking varying continuous streams of data from different sources, such as web, logs and monitoring tools can performed by a big data solution. Companies that need to capture, store and access, but not necessarily process data in real-time, can turn to big data.

Right Tools	Right Intent to execute	Right Vision	Resulting Solution

Generally, a well-executed traditional analytics solution with the right tools, right execution and a strong vision is sufficient for most companies. So, big data tools should be considered once their existing solution is not longer able to effectively store and processes valuable data.

Right Tools	Right Intent to execute	Right Vision	Solution No Longer Effective	New Solution using New Tools

Is big data a problem worth solving?

Once a company has established that they do have a big data problem, the next question to ask is "is it worth it to solve my big data problem?" Many of our clients are looking for ways to reduce costs associated with storing and processing data, and look to big data tools as a possible answer. Partly because tools like Hadoop and NoSQL datastores are open source with

Exploring The Misconceptions

There are a number of common misconceptions about big data, created in part, by marketing campaigns and media hype. Like, for example, that big data is synonymous for "a lot of data", or only the "Volume" aspect of big data. The reality is that companies like IBM & Teradata have been creating solutions for dealing with large data volumes for over 30 years. When choosing a solution to solve a big data problem, all "Three V's" should be taken into consideration.

It's important to keep in mind that big data solutions are not meant to replace existing technologies. While there is some overlap in capabilities between existing solutions and the new open source big data tools, they are meant to complement existing architectures to solve a new kind of problem . It's also important to note that a big data solution is not stand-alone. Unless it can integrate into the rest of your analytics and application stacks, its value will be severely limited. Some of our clients have made the mistake of using the term big data synonymously for a single technology. This is another misconception: big data problems usually cannot be solved with a single technology, such as Hadoop.

Another popular misconception is that big data technologies can perform analytics in real-time. This is a common misunderstanding of "velocity" of data. The reality is that while these tools can easily scale to support capturing large volumes of data in real-time, analyzing the data in a meaningful way (beyond simple counters or with limited data) can actually be quite slow. The general thinking is that data analysis queries that used to take many hours or days can be completed in about 30 minutes to a few hours.
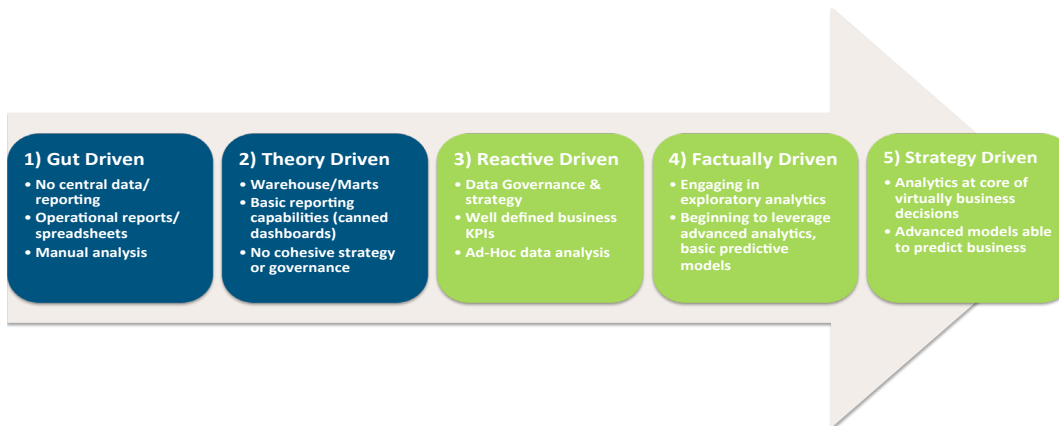
relatively cheap commercial support. But in reality, building and maintaining big data solutions is expensive.

Big data tools like Hadoop require maintaining clusters of servers and hiring developers to leverage their platforms. They are also usually command line driven which requires the end user to understand things like the Linux command line.   So while some money might be saved on licensing costs, there's still hardware, operational, development, and complexity costs to consider.

Before pursuing a big data solution, companies need to assess their analytical capabilities. If they struggle to properly analyze their non-big data assets, adding a big data tool will only lead to more complexity. Companies should be at a maturity level of at least level 3 before exploring big data tools.

In many cases, simply putting a focus on fixing or enhancing existing analytics capabilities would provide the most value.

Another question is whether or not there is value within the big data they are planning to analyze. Simply because one has big data doesn't mean the cost would outweigh the expense of analyzing it. Companies should look to execute a small proof of concept on their big data to see what insights can be determined. Examples could be as simple as manually looking through samples of log files, to leveraging a cloud-based solution, to rapidly prototype a solution based on a static set of data. With this done, an organization can also start to determine the best approach to solving their big data problem. There are many options available such as on-premise, cloud based, custom built or leveraging a packaged industry solution.

1) Gut Driven	2) Theory Driven	3) Reactive Driven	4) Factually Driven	5) Strategy Driven
• No central data/ reporting • Operational reports/ spreadsheets • Manual analysis	• Warehouse/Marts • Basic reporting capabilities (canned dashboards) • No cohesive strategy or governance	• Data Governance & strategy • Well defined business KPIs • Ad-Hoc data analysis	• Engaging in exploratory analytics • Beginning to leverage advanced analytics, basic predictive models	• Analytics at core of virtually business decisions • Advanced models able to predict business

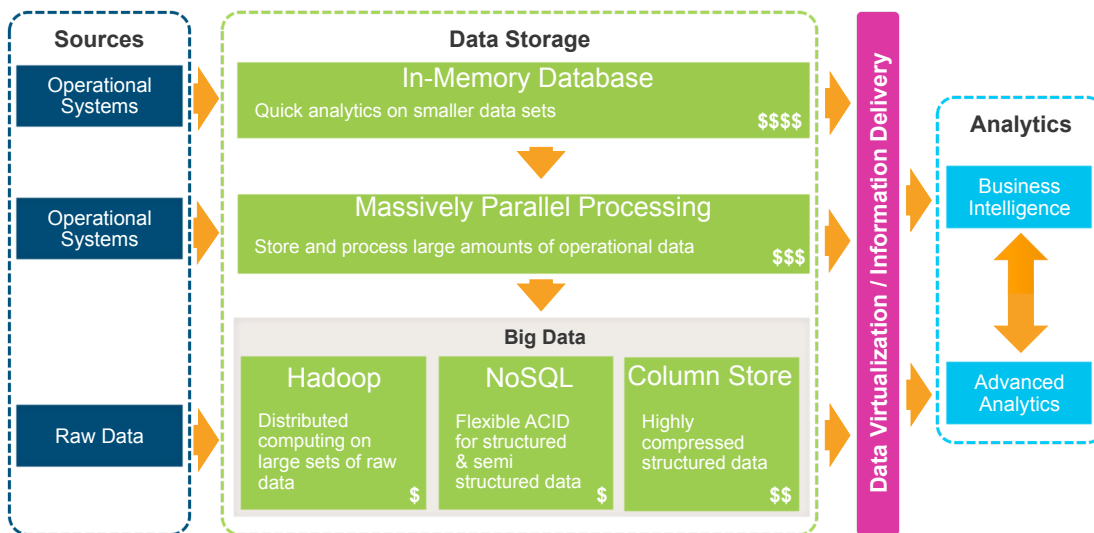| New SQL	MPP	NoSQL	Hadoop
• In Memory • Structured • Fast Analytics • Gigabytes-Terabytes	• Mixed Storage • Structured • Data Warehouses • Terabytes – low Petabytes	• Mixed Storage • Semi/Unstructured • Supports Applications • Terabytes+	• Disk Storage • Semi/Unstructured • Batch Processing • Many Petabytes+

Solving The Big Data Problem

When solving big data problems, companies need to have a full grasp of the problem at hand. There are a wide variety of data solutions available, each with their own advantages and disadvantages, however, there is no one solution that can effectively solve all problems. In order to choose the right technology, you must know what data needs to be captured, where is it going, who is using it, and why they are using it.
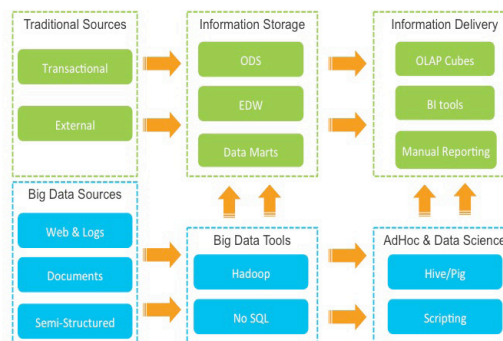
Many factors need to be taken into account when choosing a technology to solve a big data problem including cost, performance needs, and the underlying structure of the data. The companies that pioneered big data leverage a wide variety of platforms

to solve their data problems. For example, companies may use some of these tools in the following ways:
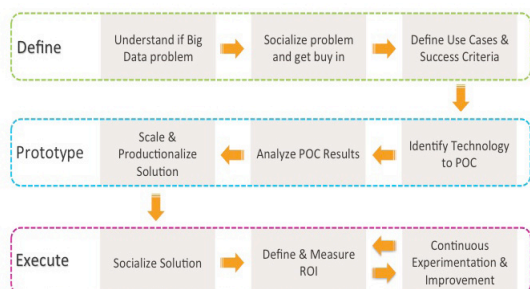
◻ A scalable in-memory solution to quickly analyze metrics based on a few days worth of data in order to identify quickly changing trends and address them.

◻ An enterprise data warehouse for analyzing big data that has been properly processed, structured, and cleansed along side traditional data for calculating KPIs, creating visualizations, etc.

◻ A NoSQL tool such as HBase to serve real time counters, basic metrics, and data signals.

◻ Hadoop for storing the most detailed log data with its entire history for both ad-hoc and batch analysis of atomic data.

Big data tools must be able to integrate into existing architectures to get maximum value. Companies that have successfully adopted big data solutions build them alongside their traditional architectures. They will typically perform ad-hoc and exploratory analytics directly on their big data platforms. Scheduled extracts are also created to get the valuable data into a reportable format and into a platform that is easier for tools and people to leverage. The most important part of analyzing big data is that you can use it along side your existing data tools. This is why virtually all companies who have adopted big data solutions use an enterprise data warehouse a downstream platform for big data analysis.

Implementing a big data platform tends to be somewhat different than traditional enterprise applications. In general, different big data tools solve a wide variety of data problems. Due to how new some of these technologies are and how rapidly they are changing, companies must be willing to experiment, fail, and try a different solution to solve their big data problems. And even once the platform is live, there should be continuous development and experimentation. The area of big data is still very new and each company's data problem is unique. A high level approach to big data might look like the following:

Define	Understand if Big Data problem	➡	Socialize problem and get buy in	➡	Define Use Cases & Success Criteria

Prototype	Scale & Productionalize Solution	⬅	Analyze POC Results	⬅	Identify Technology to POC

Execute	Socialize Solution	➡	Define & Measure ROI	⬅	Continuous Experimentation & Improvement

Analyzing big data can be a daunting task for companies who are just beginning to approach it. If not focused on the right things, it can quickly spiral out of control. By asking the right questions looking beyond the hype and marketing, companies can determine if analyzing big data is right for them.

About Slalom Consulting

Slalom Consulting brings together business and technology expertise to help companies drive enterprise performance, accelerate innovation, enhance the customer experience, and increase employee productivity. The firm delivers award-winning solutions in areas such as information management and analytics, sales and marketing, organizational effectiveness, CFO advisory, mobility, and cloud through a national network of local offices and major alliance partners, including Microsoft, Salesforce.com, and Amazon Web Services.

Founded in 2001 and based in Seattle, WA, Slalom has organically grown to more than 2,200 consultants. The company has been ranked as a Top 10 Best Firms to Work For by *Consulting* magazine four times, and earned recognition from Microsoft as a Partner of the Year five times. For more information, visit slalom.com.

Chris Schrader
Solution Principal
chriss@slalom.com