WIPRO
Applying Thought

How to find Big Value in Big Data and Analytics

The chaos and confusion around data

KR Sanjiv
Senior Vice President and Global Head - Analytics and Information Management

Table of contents

The Chaos and Confusion Around Data

Who or what today isn't producing data? Raise your gaze and there is a metric to be captured! Businesses, customers, equipment, machines, sensors, gauges, devices, systems, processes, documents, spreadsheets, customer databases, social feeds, instant messaging, mobile interaction, anything and everything is spewing out an unusual amount of data. It is a flood of numbers. Businesses are overwhelmed by the sheer volume, velocity and complexity of the data. It's ironic that in a world of shrinking resources the one plentiful thing we have – data – is threatening to smother us.

Then there are those who are sitting up and saying, "Let's get to the bottom of this one and see how we can dig out all those business nuggets and insights buried within the data. Let's turn this data into an asset." For instance, an oil rig generates about 30,000 data points per second. Less than 5% of that information is used. Oil rig bosses are waking up and saying, "I want to improve my productivity and efficiency leveraging historical data and real time machine generated data to save millions of dollars." Businesses like these have begun to capture and consume an inordinate amount of the data to become more competitive, uncover fresh trends, drive innovation, design new products and services, create efficiencies, reduce costs and improve decision making.

Today, many businesses recognize the need to roll up their sleeves and get their hands dirty with data. But they are uncertain of where to begin. The resources required, the tool kit that will help sift through the chaos of numbers, the people, pitfalls and the costs involved are not clear. What they do know is that Big Data and Analytics are going to revolutionize performance management and customer insight, altering forever the way we view quality and market responsiveness.

Big Data and Analytics are at the center of action. They are unlocking Big Value when businesses identify the right use case.

How Others are Using Data and Analytics – Are You?

Here are two quick examples to illustrate how data and analytics are uncovering new businesses, building new revenue streams and adding to bottom lines:

- A manufacturer of heavy mining equipment has begun to charge for drilling using per kilometre pricing against stringent SLAs. This means monitoring assets in real time, predicting failure and ensuring equipment is serviced before it breaks down. Leveraging data and analytics in real time has helped the company create an alternative revenue stream for its business.

- A leading manufacturer of surveillance equipment started collecting data from its devices installed at various customer locations. Using the vast amount of data at its disposal, it could create models and metrics that help in surveillance and fraud management. The manufacturer began to offer solutions in the area of surveillance and fraud management, creating a completely new revenue stream.

To do this, businesses need to go well beyond the customary approaches to data that utilize databases, data warehouses and data marts. These last-century data management systems cannot cope with high volumes of data running into petabytes. They cannot process the data in real time for quick decisions. They are unable to handle unstructured data from sources such as social media conversations, leaving large gaps in understanding what is happening out there to business.

The new approaches require adept teams to manage Big Data. They need appliances to handle and massage the data. In-memory computing, pattern hubs and the ability to harness data in motion are the new weapons to tame the Big Data beast.

But it's Not as Simple as it Seems

Data is not quite as simple and straightforward as we imagine it to be. It isn't always a row of numbers noted from a gauge. Sometimes it is buried in conversations and video recordings. Take a high tech computer services company that gets a customer call for technical support. The caller wants to transfer the virus protection from his laptop to the one his daughter has. "She just joined college and got a new laptop," says the father to the remote technical support agent on the phone. "The anti-virus would be useful for her." The agent deactivates the anti-virus application from one system and installs it on the other,

changing the customer credentials in the online records of the anti-virus product. The call is closed successfully. The father is delighted. His daughter's data is protected within a matter of a few minutes, without having to make a pesky trip to the product's downtown showroom. At the support center, incident logs show that SLAs were met and the call was closed successfully. Does the story end here? It may have, until a few months ago. Not any more.

The amount of data created during such calls is remarkable. In this case we know that the new customer is female and has just joined college. The old customer may want a new license of the anti-virus product shortly. The household owns two laptops of which one is new. And perhaps this is a middle-class household. This is a valuable lead in the hands of a sales person. The data, unfortunately, is teasingly buried in an audio file. If that file could be converted to text and analyzed quickly enough, bingo, another sale.

There is data rushing out of everywhere. Enterprise data – the more traditional types of company information related to finance, sales, HR, administration, legal and such – has been growing rapidly, especially in the light of regulatory requirements. This is structured data that is relatively easy to handle.

Add to this the historical data records of customer profiles, data from educational and medical records, sources such as insurance claims and driving violations, sports events, weather conditions and so on. This could largely be referred to as "data at rest" – details about people and events stored in vast data warehouses and data marts, ready to be called up for post mortems by data scientists who want to uncover lessons from the past.

Super Optimizing Process to Build Insights

Today, data is also being generated by machines that are equipped with motion sensors, cameras, microphones, GPS and accelerometers; by transactions over ATMs, mobile phones and credit cards; by social networks, sources such as Twitter and YouTube and across emails; by the clickstream of consumers on the web; field forces sending back data from mobile barcode readers, RFID sensors and countless other sources. This is data in motion. Its true value is unlocked when consumed in real time.

As an example, today's medical equipment in a hospital may produce patient temperature readings every minute. Doctors and medics will be unable to make their way through such intense data records. What they need to know is if the patient's temperature was within range during certain hours. What they need are engines that summarize the data, making it usable when the doctor arrives. They need to super optimize

analytical processes around data that is being rapidly generated and needs to be consumed just as rapidly. How do organizations get their arms around such vast quantities of rapidly moving data and correlate all of it?

There's more. What is required to manage data that is structured and unstructured? That comes in a variety of formats from audio to text, binary and video? That must be analyzed in real time as well as in batches? It is time for organizations to change their information management landscapes. What organizations were doing with gigabytes (GB) of information cannot be done when the information grows to petabytes (PB) and zettabytes (ZB). From a time perspective running batches of such huge amounts of data will take days. From a financial and an infrastructure perspective, this is not sustainable.

Organizations are finding ways to manage the data deluge using appliances and in-memory processing to ensure that users don't have to depend on the traditional methods of slicing and dicing data. Instead, this is being replaced by a discovery-based process where users can move around the data and across various dimensions of the data. When organizations develop the ability to move across dimensions of information, they begin to create new insights and better decisions. The patterns and insights created start to unveil powerful transformative and business opportunities.

Impact of Analytics

We believe that analytics will bring about major changes (see Figure 1), especially in improving customer insights; machine data that did not exist until a few years ago will assist in previously unimaginable capabilities; analytics will help in improving safety and security; and analytics will assist in the reduction of IT costs (this should bring joy to the hearts of any CIO!). Not surprisingly, leading analyst firms have named BI and analytics the No 1 CIO priority for 2012.

Improving Customer insights: Analytics can help provide a 720^0 view of the customer, driving extreme personalization (as we saw in the example of the father switching the virus product to his daughter's laptop); offer competitive pricing to select groups of customers; improve brand loyalty; and drive consumption through targeted offers and campaigns.

Next-generation interaction solutions are being built by retailers that integrate data from different interaction points such as social media, web logs, store wi-fi, past purchase history, CRM, etc, to create real-time next best offers to improve customer engagement. The offers are delivered through various channels like store kiosks, web and mobile, based on analytics that determine the most likely channel to produce a "buy" decision at that point of time.
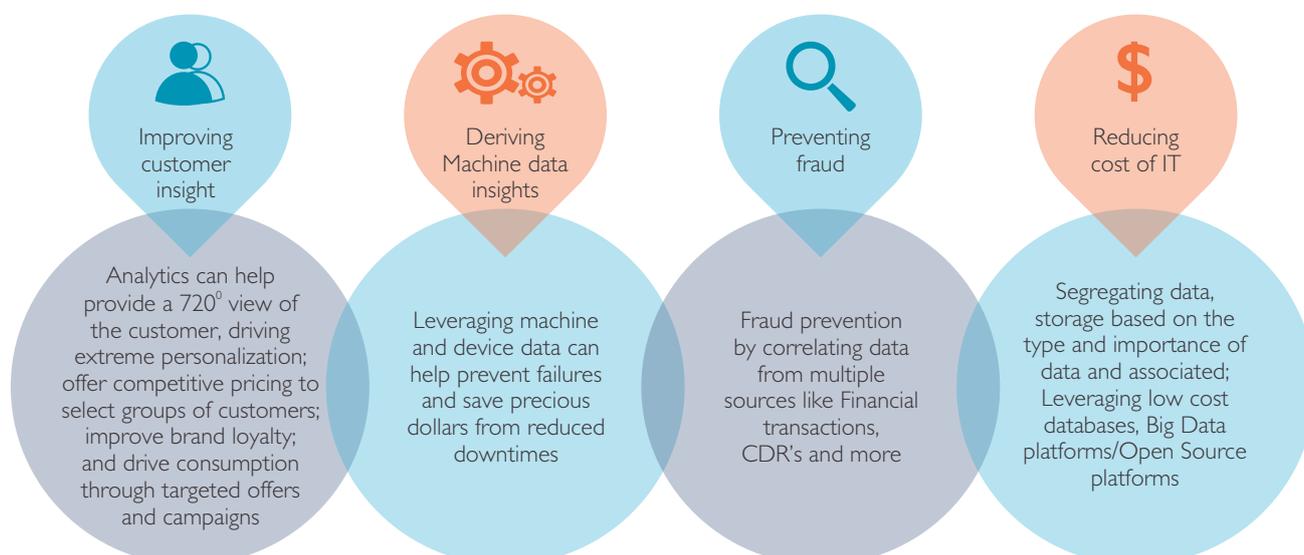
Deriving Machine and device data insights: Leveraging machine and device data – such as those in oil rigs and mines, data thrown up by digital TV set top boxes and smart cards etc – can help prevent failures and save precious dollars from reduced downtimes (we saw a slightly more critical outcome in the form of doctors being able to make better decisions for patients thanks to machine data and analytics).

A single airplane engine generates more than 10 TB of data every 30 minutes. In the past, this data was deleted at the end of each flight. Now the same data can be used to monitor the health of the engines and replace them before they fail. Data can also be used to shape engine technology and create green engines with lower fuel consumption & lower noise levels.

An insurance company collects telematics data from drivers who have signed up with them and offers pay-as-you-drive insurance schemes based on driving patterns.

Preventing fraud: Machines can help identify and mitigate fraud and improve surveillance capabilities. There are a higher number of data points from newer data sources (telecom systems, GPS, social media, etc)

Figure 1: Major changes on the way, thanks to analytics

Improving customer insight	Deriving Machine data insights	Preventing fraud	Reducing cost of IT
Analytics can help provide a 720^0 view of the customer, driving extreme personalization; offer competitive pricing to select groups of customers; improve brand loyalty; and drive consumption through targeted offers and campaigns	Leveraging machine and device data can help prevent failures and save precious dollars from reduced downtimes	Fraud prevention by correlating data from multiple sources like Financial transactions, CDR's and more	Segregating data, storage based on the type and importance of data and associated; Leveraging low cost databases, Big Data platforms/Open Source platforms

available to create and store pattern repositories that can help expose fraud faster and reduce false positives.

In a recent case, a lady in Canada claimed insurance for disability. The fraud in this instance was uncovered when the company found pictures of the lady on a skiing holiday posted on Facebook.

Using machine data and newer data sources, rogue trading can be avoided by correlating data on trade logs with portfolio & asset value, risk rating, P&L statements etc. A typical customer has over 250+ databases that store trade, portfolio, risk rating related data, etc. Correlating this to identify trading patterns takes days and sometimes months. Big Data provides the ideal platform to correlate volumes of data and process.

Reducing cost of IT: Segregating data and assigning storage based on the type and importance of data and associated processing is one way of bringing costs down. We don't need expensive databases. We can leverage low cost databases, Big Data platforms and Open Source platforms. Leveraging low cost Big Data platforms for storage of raw and unprocessed data and moving only pre-processed data to the EDW allows reducing volumes on the EDW. Costs can also be controlled through enhanced low-latency operations like real time data loads and queries through the use of No-SQL Databases.

Don't Take a Leap of Faith – Try Being Rational

It really boils down to asking the key question, ''Where do we begin the data and analytics journey?'' Rather than throw up a mountain of internal research and justification for adoption, we suggest using a simple framework to undertake the journey. This is what we call the 3B Framework for Adoption:

Build a business case: If you don't have a business case, you don't need to embrace Big Data and Analytics.

- Evaluate revenue generation models and cost optimization models

- Identify business processes that will provide a Quantum differentiation

- Develop models for ROI and Pay-back period

Base-lined Data Processes: Get the fundamentals in place before you begin so that the journey is smooth and uninterrupted.

- Identify and leverage internal data sources

- Identify external and new age data sources to plug gaps

- Baseline Data Quality and institutionalize Data Governance

- Ensure strict adherence to Data Privacy rules and regulations

Best of breed technology: Tools are what will bring in results. Ensure the best.

- Identify gaps in the current technology landscape

- Identify technologies for data processing and storage, ETL/ELT, visualization & predictive analytics

- Carry POTs and select technologies

Adopting Big Data and Analytics to power your business is inevitable. Finding the right way to do it is the only challenge.

About the Author

K. R. Sanjiv is the Senior Vice President and Global Head for Analytics & Information Management, Wipro Technologies. He carries P&L responsibility, strategy and operations of this unit globally reporting to CEO.

Analytics & Information Management helps customers derive valuable insights out of integrated information by bringing together the combined expertise of Analytics, Business Intelligence, Performance Management and Information Management.

The group provides consulting, business centric and technology specific analytical solutions and data management frameworks developed through a complete ecosystem of partners, focusing on industry specific analytics, optimization and operations analytics, Enterprise Data Warehouse, MDM, Data quality and data life cycle management.

Sanjiv has over 20 years of enterprise IT and security experience, including consulting, application and technology development that spans multiple industry segments and diverse technology areas.

Since joining Wipro in 1989, Sanjiv has been involved in defining enterprise architectures for organizations that included technical models, transformation program definitions and governance models. He has designed OLTP mission critical systems such as screen-based trading systems for stock exchanges, surveillance systems and order routing systems for brokerage houses. He has spearheaded due diligence exercises in M&A situations for customers and also managed large project implementations in a global delivery model.

Sanjiv has spoken at leading CXO summits, industry and academic conferences on varied topics related to Business Technology. He holds a bachelor's degree from Birla Institute of Technology and Science, Pilani.

About Wipro Technologies

Wipro Technologies, the global IT business of Wipro Limited (NYSE:WIT) is a leading Information Technology, Consulting and Outsourcing company, that delivers solutions to enable its clients do business better. Wipro Technologies delivers winning business outcomes through its deep industry experience and a 360 degree view of "Business through Technology" – helping clients create successful and adaptive businesses. A company recognized globally for its comprehensive portfolio of services, a practitioner's approach to delivering innovation and an organization wide commitment to sustainability, Wipro Technologies has 135,000 employees and clients across 54 countries. For more information, please visit www.wipro.com

WIPRO
Applying Thought

DO BUSINESS BETTER

NYSE:WIT | OVER 135,000 EMPLOYEES | 54 COUNTRIES   CONSULTING | SYSTEM INTEGRATION | OUTSOURCING