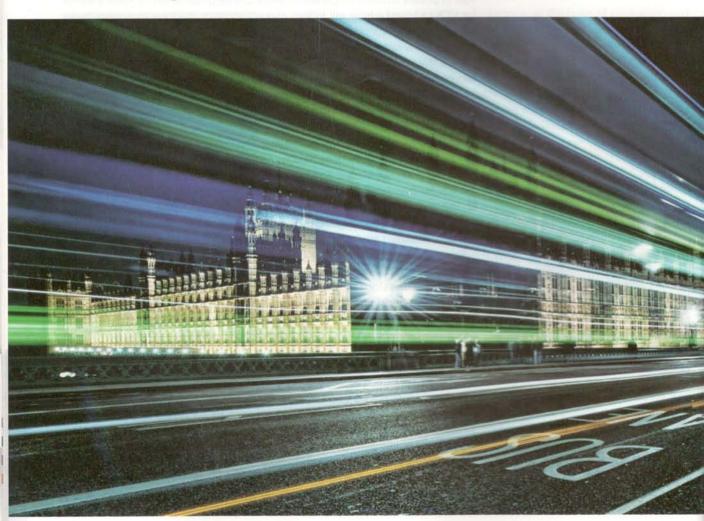Onkar Sharma
onkars@cybermedia.co.in

Demystifying Big Data

Implementing big data is one of the buzzwords in the enterprise IT world today. Often the debate around this overshadows the argument whether it is required or not. There are many sides to big data which most CIOs need to know, if big data is in their hands or not

Big data is one of the promising concepts that stands to change the way businesses behave and look at customer service. The ability to analyze their data precisely and with clinical precision makes them competitive beyond imagination. With the lightening fast speed, they can sense what customers desire and when. In simple words, it opens up a Pandora's box of opportunities for them. Realizing the power of big data, most of the companies want to embrace them at the earliest. The numbers also speak of this. Globally the Big Data analytics and related technology market is predicted by IDC to grow at a 26.4% compounded annual growth rate and is expected to reach US$41.5 billion by 2018. In fact by 2020 IDC believes that big data will be one of the key drivers for the economic growth of any nation worldwide.

According to Nasscom, the Indian market is still in early stages of adoption of big data analytics. However, with surplus talent, established infrastructure, and a mature ecosystem, India is on its way to become a global hub for analytics. In India, there are verticals like banking, telecom, etail and, to an extent, government, where big data has begun to unleash its power. For instance, Ecommerce, which is witnessing the fierce competition, seems to have taken the early plunge. Today every etailer is elbowing through a horde of competitors to steal the business. Big data and analytics are helping them to read consumer minds in the best possible way.

Without harping on the benefits of big data, it is critical to take into cognizant the realistic view of how-to- and who-to-use big data. Big data is a fancy word in today's enterprise IT world which every vendor wants to exploit. As a customer or CIO, it is better the implications of the technology are taken into consideration. You need to ask a few questions.
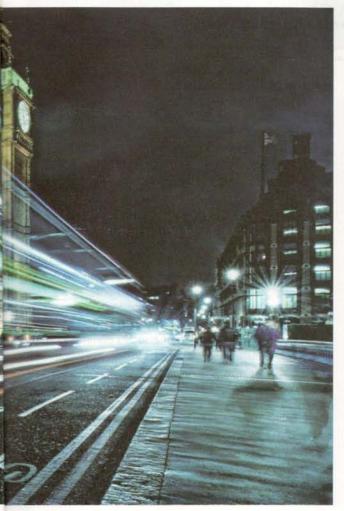
Is your organization creating humungous amounts of data? Are you able to manage the flow of data? Are you using data to take informed business decision? How quickly are you able to extract the insights?

INSIGHTS INTO OPERATIONS

Embedding analytics into operations is one of the biggest challenges CIOs face today. Anuj Kumar, GM India and South Asia and Global Lead – Analytics, Concentrix underlines this fact when he says, "The process of bringing data-science to work closely with business is an evolving art which is yet to be fully mastered. The objective of big data analytics is to bring out meaningful insights and not just present data more smartly. This is best done by working closely with business, and through an act of decluttering and simplification." Data is the best way to know how operations help the companies to steer forward.

Experts believe that large organizations across industries are now learning to deal with humungous amounts of data generated in the course of their everyday operations and through customer interactions across channels, including social media. While this data is being used currently to generate insights to a certain extent, organizations are faced with the challenge of embedding analytics in their operations.

In one of his earlier addresses R Chandrashekhar, President, Nasscom had said, "Big data offers a unique suite of advanced analytics and helps derive meaningful insights from customer data to increase sales, better target customers, improve reach and gain competitive advantage. The Indian market is still in early stages of adoption of analytics."

Interesting case of Bangalore Water Suppy and Sewerage Board

Bangalore Water Supply and Sewerage Board (BWSSB), which supplies water to nearly 800 square kilometers of the city, is using big data and predictive analytics technology to create systems for monitoring and managing increasingly complex water distribution systems.

Bangalore's massive population growth—from 5.4 mn in 2000 to over 10 mn and counting today—has put tremendous strain on the city's water supply and distribution systems. The main sources of water(Cauvery and Arkavathi rivers) are not just sufficient to meet the water demand in the city to a permissible per capita norms. This leads to a big challenge in equitable distribution of available water across the divisions/subdivisions. BWSSB created an operational dashboard, based on the IBM Intelligent Operations Center (IOC), which serves as a 'command center' for monitoring, administering and managing the city's water supply networks.

The command center monitors the waterflow in 284 of 784 bulk flow meters in the city and provides a clear, single view of the functioning of all the bulk flow meter, amount of water transmitted by each of them, the amount of water supplied to individual parts of the distribution system, the level of water in each reservoir or tank etc. Data from every working flow meter will be reported on a single dashboard. The Intelligent Operations Center based solution contains the GIS (Geo Information System) for Bangalore to enable a real-time view of flow meters, along with the ability to zoom in and out, and pan and click on a specific flow meter. When an asset (GLR or flow meter) is selected, a user can have a view of the Key Performance Indicators (KPIs) such as latest flow rate, total flow in 24 hrs and average total flow over past seven days, as well as the geographical location and time of last update.

Dr P.N Ravindra, Executive Engineer (New Initiatives and Design Cell), BWSSB was quoted in a case study by IBM saying, "Around 45% of the water supplied by the BWSSB goes unaccounted and implementing this solution helps minimize unaccounted for water by detecting large changes in water flow, through real-time monitoring."

BWSSB engineers can now make modifications in the settings of the control valves and get real time feedback on the changes to the water supply elicited by their actions. By setting and adjusting thresholds at key points, engineers can ensure that supply meets their expected goals. When these goals are not met, real time alarms will allow the engineers to make quick, informed decisions and modifications.

WHO NEEDS BIG DATA: OPTIONS VS SOLUTIONS

It is not difficult to know whether or not big data is suitable for an organization. If the volumes of data being generated on a daily-basis is extremely high which can not be handled using traditional analytics tools, it is time that your organization need to consider big data. In addition, taking insights into the businesses where big data is being used helps in understanding and implications.

Many believe that there has to be a price-demand balance for using big data. "As an example, Big data or large amounts of data is useful only when it enhances the ability of the analytics based Revenue management system (RMS) to provide better price-demand estimates, thus enhancing the optimization process," suggests KS Prashant, Managing Director, IDeaS Revenue Solutions

Big data is much beyond the volumes. The volumes which we generally believe can only be handled by big data may require an alternate tool. Here is how. "In general Big Data makes sense for scenarios where the solution needs to solve all the 3 Vs – Volume, Velocity and Variety," says Sreejit Menon, Program Director - Information Management, Happiest Minds Technologies. "If it's just volume, databases like SQL Server, Oracle, etc. can be migrated to powerful object-relational database systems like PostgreSQL which can support unlimited database size including 32 TB table size, 1.6 TB Row Size and 1 GB Field sizes with unlimited indexes per table. The other

option would be to move to appliances like IBM Netezza, SAP HANA, EMC Greenplum, etc," he suggests.

Besides, there are Enterprise Application Integration (EAI) solutions that have been available for over a few decades now that are real-time / near real-time and the leaders in this segment are TIBCO, IBM, Oracle, etc. Event series analysis and event correlation engine or complex event processing engines like Esper are a popular open source alternative.

Menon further outlines that to deal with variety, there are content management solutions that work with variety of data including documents, audio and video, etc. Techniques like Natural Language Processing (NLP) using enterprise Text Mining software from SAS, SPSS, etc. and open source software like Python Natural Language Tool Kit or Apache OpenNLP or Stanford CoreNLP enable analyzing Text to a large extent and are commonly used to extract features and derive sentiments. Unstructured Information Management Architecture (UIMA) is a popular framework for Content Analytics. Making another strking point, Menon believes, "It is in solving all the problems together that Big Data Analytics platforms make a lot of sense - typically with volumes greater than at least 1 TB of both transactional as well as digital (social + others) and need for real and near real-time analytics to start with."

Virender Jeet, Senior Vice President, Newgen Software says, "Big data holds relevance in scenarios where the or-

ganization has basic infrastructure to store voluminous data. Also, Big Data makes more sense when the data is around customers that can be analyzed to define business strategies and predictive modelling of consumer behaviour."

TIME TO HARNESSING DATA

In big data analytics, data sufficiency plays a crucial role when samples are run across different dimensions. Sufficient data points to make adequate analyses are required. Not only the quantity of data, the quality of data being used for crunching, too, influences the quality of insights.

According to iDeaS, there are some data sources that can do more harm than good. As an example, weather and airline data is measured frequently and widely, and these are very broad measurements. Weather and airline data trends may be impactful to travel patterns at large, but their relationship to business or leisure bookings in a particular location is loosely coupled. Further, in the case of weather data, much of this data itself is forecast, thereby introducing yet another possible source of error into the demand forecast model.

In short, enterprises need to go beyond their organizational data to fetch accurate insights. Since weather has direct impact on airlines, airlines need to understand the weather trends and assimilate them into their data. "It is not enough just to have large amounts of data available. When solving for competitive advantage, enterprises need to glean insights from the data, in addition to layering in the right data relative to their market. They should possess the ability to be able to process the data, address business problems such as revenue or profit optimization using advanced analytics and be able to deliver the results to practitioners in an easily consumable manner by using the relevant technology platforms and hardware with huge data processing power," opines Prashant from iDeaS Revenue Solutions. This kind of domain, analytics and technology along with the right hardware and software investment mix is required to solution for competitive advantage and often times that becomes difficult to piece together.

NEED OF CHIEF DATA OFFICER

Many enterprises still don't consider data as a strategic asset and, hence do not have a clear strategy for monetizing their data. At times, the amount of data that these organizations generate or have present over years in multiple siloed stores is humungous. "There needs to be an enterprise-wide data strategy initiative supported or anchored along with key Business Stakeholders for developing a roadmap and execution plan for the same, including MDM or and Data Governance programs to enable the larger objectives. Many organizations hence, have roles like the Chief Data Officer," advises Menon from Happiest Minds. He further asserts that it shouldn't be confused with Chief Digital Officer whose role would at times have an overlap, but is clearly set for a different objective, more towards enhancing customer experience and improving operational efficiency.

CONFUSED BETWEEN ANALYTICS AND BIG DATA

Technically speaking, big data is the Information Management layer and provides the platform for Analytics to be performed, explains an expert. Quite a few big data Visualization and Analytics platforms and products, both on premise as well as on cloud, are now available with custom industry solutions out-of-the-box. Many are still frameworks and can be used to accelerate the implementation at times. "However, big data Analytics is mostly an initiative that combines implementation of big data (Appliance/Hadoop Ecosystem/NoSQL solutions) and, thereafter Visualization and Analytics on top of it. There are still a lot of traditional analytics being done and consumed that has nothing to do with big data as such," elaborates Menon.

While the basic techniques and applications of Analytics, be it descriptive, predictive or prescriptive, more or less remains the same even on big data, traditional Analytics used technologies like SAS and SPSS. But big data is much more than analytics. It needs more scalable and compatible analytics algorithm and involves code development in various platforms like Java, Python, Scala with use of Machine Learning libraries.

CREATING A ROADMAP

To make big data work for an organization and ensure that the RoI on the implementation is met, increased measurable insights need be generated and enabled through data science/advanced analytics. This should be on top of the big data platform with larger objectives of improving customer experience impacting positive revenue growth and/or improved business efficiency. Verticals like retail, banking, telecom and government reserve a huge potential for big data usage in India.

Siddharth Idnani, Sales Director, Oracle India says, "While many organizations have achieved proficiency in exploiting their data through data analysis, they are still at the early stages of creating an analytics model that can deliver real business value from big data. The amount of data that is being generated is increasing steadily, and businesses are still faced challenges with respect to complexity, security and costs. In order to overcome these challenges, it is critical for businesses to invest in the right technology."