

THE BURGEONING NEED FOR BIG DATA ASSURANCE



Contents



03	Executive Summary
04	Introduction
04	The State of Big Data Implementations Today
06	What is Big Data Assurance ?
06	Current State of Assurance in Big Data Implementations
07	Benefits from Implementing a Holistic Big Data Assurance Strategy
07	Conclusion



Executive Summary

Over two-third of enterprises, across industry verticals, who are currently planning or implementing Big Data analytics projects stated that their projects have not progressed past the pilot stage. One of the primary reasons identified for this lack of success appears to be the lack of a holistic Big Data Assurance Strategy to address issues related to data quality and rein in the cost of quality and delivery on projects.

Big Data has been a key focus area for enterprises in the information technology landscape for many years now. Industry analysts across the globe agree that Big Data is a multi-billion dollar industry. However, despite the buzz and spending on implementing Big Data technologies, only a few understand what the term really means. Moreover, there is a lot of uncertainty around implementing Big Data projects globally across industries due to the lack of credible market intelligence in the subject. To find out more, Wipro conducted a market research study to gain insights into the state of current enterprise Big Data projects.

Our survey of senior executives and leading analysts based in organizations in the US, Canada, Europe, and Asia – captures a vivid snapshot of the various drivers and flavors of the current enterprise Big Data project implementations across a few industries. Our research findings also provide a strategic outlook into some of the critical elements in the market such as the most favored Big Data tools and technologies, investment interests of enterprises, hurdles in current project implementations, as well as the gaps that exist in today's Big Data ecosystem.

A quick synopsis of the survey results show that:

- **Need for business intelligence drives the demand for analytics** – A majority of respondents (92%) stated that they need Big Data implementation for business intelligence (92%) and predictive analytics (88%). Companies across all industries, except utilities, unanimously stated business intelligence as the primary use of Big Data. In Utilities, fraud detection topped the list of needs (100%) while predictive analytics ranked second (80%).

- **Big Data analysis (69%) and Assurance (52%)** were identified as the key service/process-related gaps in current implementations – around 77% of respondents said it's difficult to find Big Data Assurance skills in the market. To bridge this gap, 12% of respondents are already working with third-party vendors to meet their Assurance needs and 42% see this option growing significantly in the next three to five years.

- **Standard solutions are most preferred to solve current Big Data needs** – 81% respondents stated that they prefer to work with standard solutions to solve their Big Data needs. Enhancing in-house talent pool (79%) and working with a third-party to define a process or solution (75%) came close on the heels, while surprisingly, over a half (54%) said they will invest in developing their own Big Data solutions.

- **Hadoop technologies leading tech adoption** - 80% of our respondents stated that they primarily use Hadoop technologies to store and process their data. Cloudera (32%) appears to be most popular Hadoop platform, followed by Hortonworks (17%) and MapR (17%). Other than these, Splunk at 20% is the other tool of choice.

- **Moving past pilot stages is proving to be a tough task** - Although over two-third of the organizations interviewed having Big Data implementations planned/underway over the next five years, very few projects have progressed past the pilot stage. Moreover, these implementations primarily tend to be on limited datasets, which indicate that most organizations have low levels of maturity in terms of data integration and data processing using Big Data technologies.



Introduction

Fueled by plummeting data storage costs, Big Data technologies have grown exponentially over the past decade. Big Data platforms, which started out as batch-processing platforms, are now evolving to handle near real-time processing. These platforms allow continuous access and processing of data at a large scale to aid datafication – the ability to discover previously unknown/unseen trends and relationships using data. This trend has further accelerated with the explosion in the number of connected devices collecting data. As data collection increases in scale and scope, there will be a larger need to assess data processing agility and data quality. New processes and policies to measure the speed and efficiency of capturing, managing, and analyzing data will come into effect, which in turn will help consolidate standards and policies for data management and analysis. In order to support this trend, there will be a rise in the number of self-servicing Big Data applications and platforms based on PaaS (Platform as a Service) and SaaS (Software as a Service) models.

As a result of, the exponential generation of data, managing the volume, velocity and variety is posing a herculean challenge to businesses

worldwide – which has resulted in issues related to data quality. Gartner estimates that poor quality of data costs an average organization \$13.5 million per year, and yet data governance problems — which all organizations suffer from — are worsening.¹ Wipro believes that the key reasons for poor data quality include:

- Deluge of structured and unstructured data and its growing complexity
- Siloes of data storage within an organization
- Changing technology landscape and rising integration complexities resulting in inconsistency of data,
- Lack of a common/standard language in business information management

Can Big Data Assurance help minimize losses associated with bad data quality? To answer that question, let's look at some of the current Big Data needs of organizations.

The State of Big Data Implementations Today

We found that most customers today appear to be taking their first steps towards creating a data processing engine for their various data processing needs. This is evidenced by the fact that 80% of our respondents stated that their biggest spend on Big Data technologies were on the three leading Hadoop platforms (Cloudera - 32%, Hortonworks – 17%, MapR – 17%, and the base version – Apache Hadoop – 10%). After setting up their platforms to store and archive data – typically called Data Lakes – organizations seek to create data models that allow them to derive value from their data. This process of value creation from data requires the creation of data models.

Data models typically govern the following processes for organizations across industries:

- What data needs to be correlated, mined, analyzed, and, processed
- What sources provide the data for the step above
- How the data should be collected (data cleansing/scrubbing are typically elucidated) and used – also known as Data Governance
- How data processed should be consumed

¹Gartner Press Release, "Gartner Says CIOs and CDOs Must 'Digitally Remaster' Their Organizations", February 2, 2015, <http://www.gartner.com/newsroom/id/2975018>

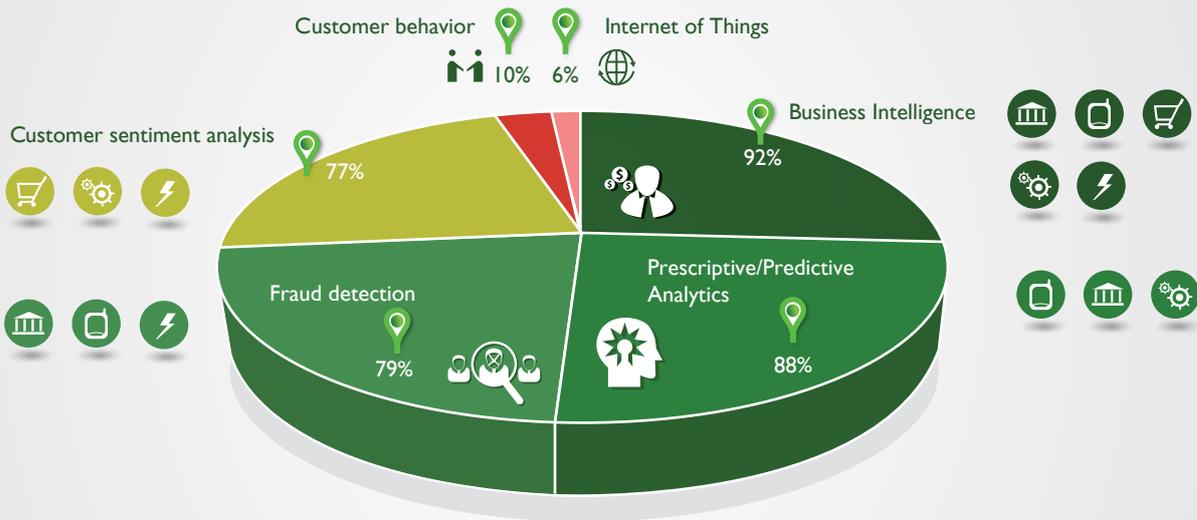


Figure I – How Big Data will be used in enterprises by industry ?

Once organizations create a data model, they are able to proceed into the next phase of their Big Data projects – Data Processing to provide value. Our survey finds out that enterprises are primarily adopting Big Data technologies and solutions to pursue projects that will provide business intelligence (92%), predictive/prescriptive analytics (88%), fraud detection (79%), and customer behavior and customer sentiment analysis (77%).

To address their technological needs, most organizations appear to be using standard solutions (81%) rather than working on creating

custom solutions/services in-house. A large part of this may have to do with the fact that organizations are currently focused on enhancing their talent pools (79%), and still defining processes and solutions by working with third parties (75%). The survey results also indicate that only a smaller percentage of organizations have reached a point where they have proceeded past the data integration phase into the data processing phase as there are no data processing tools other than Splunk (20%) that figured in the top Big Data spends for enterprises.

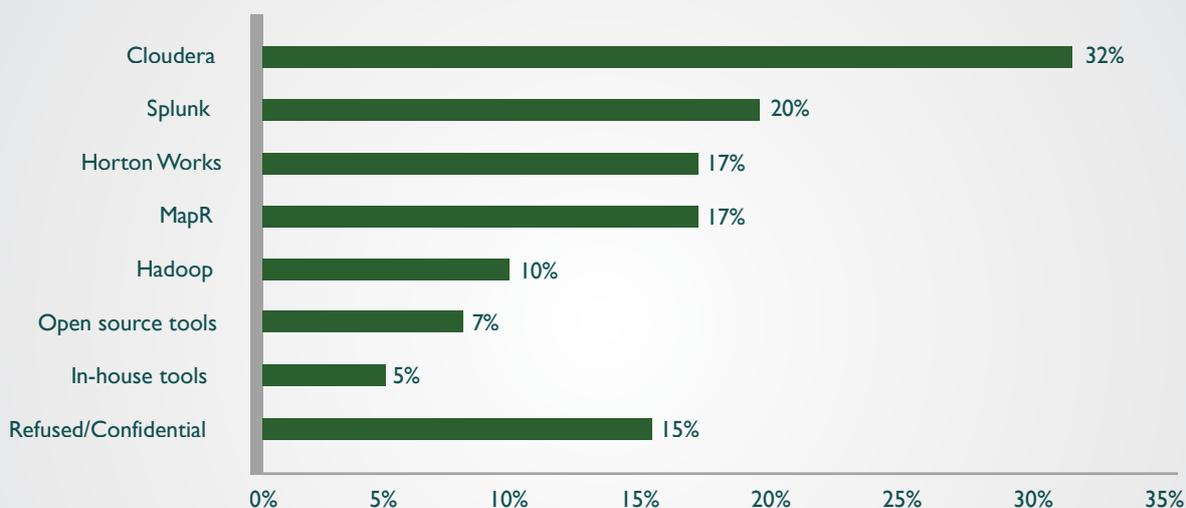


Figure II- What are the primary tools used in Big Data implementation?

What is Big Data Assurance?



Big Data Assurance is about providing a strategy, deriving a process, and aligning the right tools and resources required to address the problem areas outlined above. However, in order to create a Big Data Assurance strategy, it is not only important to understand the pain points observed in current implementations, it is also critical to understand the nature of Big Data implementations seen across enterprises. We have observed that there are currently two primary flavors of enterprise implementations that handle Big Data:

- High volume, velocity, and variety of data repositories
- High performance data processing engines

Each flavor of implementation requires a different Assurance strategy to address the issues that will be faced. For Big Data platforms used as data repositories (such as Data Lakes), the primary area of concern relates to the correctness, completeness, and timeliness of the data stored from various sources. To ensure this, two primary tasks need to be performed – first, ensure that each data source provides data that is correct and complete (compared to the source),

and second, ensure that the quality of data on the system meets the standards that meets governance and timeliness policies required.

Assuring quality on high performance Hadoop platforms require a slightly different approach in addition to the tasks associated with the first type of implementation. In order to assure quality on this type of implementation, it is not only critical to ensure the quality and correctness of data that is stored on the data system, it is imperative to test (functional and non-functional such as performance) the various algorithms that are written to cleanse, process, and transform the data that will ultimately provide the metrics, dashboards, reports, and, other consumables required.

Moreover, given that Assurance tasks in Big Data implementations involve working with large and varied amounts of data, it is imperative to have an automation strategy to ensure that resources don't spend too much time and effort performing mundane, yet, critical tasks. Despite this, most implementations today primarily perform all testing tasks in the Big Data world manually.

Current State of Assurance in Big Data Implementations

The level of maturity in current Big Data implementations show that there is a huge scope for Data Wellness/Data Assurance in current Big Data implementations. This is further proven by the fact that over 52% of respondents cited that the market currently lacks good Assurance tools and services. Only the lack of data analysis tools and services (62%) was featured as a bigger service/tool-related gap for enterprises. The demand for professionals with Big Data Assurance experience coupled with the lack of tools and services has resulted in over 77% of the respondents to the survey struggling to find individuals who have the required skills to perform Assurance using Big Data technologies. Across verticals, Utilities (40%) have taken the lead in hiring third parties for Big Data Assurance, whereas Hi-Tech (67%), Banking and Insurance (50%) companies display a high preference in keeping Assurance activities in-house.

Results also indicate that enterprises are on the lookout for testers who have expertise in data analysis (88%), Hadoop development

(58%), and tool-specific work experience (52%). The profile of individuals who will perform the role of testers is skewed towards those who can develop as well as test. Although only 29% of our respondents stated that all Big Data Assurance will be tightly coupled with development, 49% believed that there will be a role for testers who have development and scripting skills as well. Only 22% of our respondents believe that there will be standalone Assurance opportunities on Big Data projects.

Enterprises today appear to be in splits about how to fulfill the assurance resource requirements on projects. Only 29% of our respondents are either currently working with third-party service providers or considering an engagement with third-party providers whereas, 32% of our respondents expect to hire professionals outright from the market. Almost 39% of our respondents are unsure about what strategy they need to adopt at the moment.

Although organizations are not planning to look for people resources for their Big Data projects, it appears that enterprises will look at engaging third-party tools and consulting services. This can be inferred by the fact that 80% of our respondents will engage with third parties for their Big Data Assurance requirements over the next

three to five years. Overall, companies with budgets between \$50M and \$100M are most likely to outsource Big Data Assurance to third-party vendors. This shows that there is a lot of scope for Big Data Assurance tools and services (especially Big Data Assurance as a service) in the near term for third-party vendors.

Benefits from Implementing a Holistic Big Data Assurance Strategy

From the results mentioned above, it is clear that organizations are unable to implement a holistic Big Data strategy. This situation could have arisen due to a lack of understanding on how to integrate an Assurance strategy into current implementations, along with a paucity of skilled resources, services, and tools available to address the specific challenges posed on current Big Data implementations.

The best method to integrate a Big Data Assurance strategy – into existing implementations – would be to understand the essentials of what it involves, and how it will help organizations overcome current

challenges. A Big Data Assurance strategy helps enterprises derive maximum value from their Big Data implementations. A holistic strategy should primarily deal with:

- Testing the quality of the data right from the source through the use of a data quality index for each data source/stream
- Testing the quality of the platform and the algorithms/technologies used to transform the data from its raw state to a consumable state

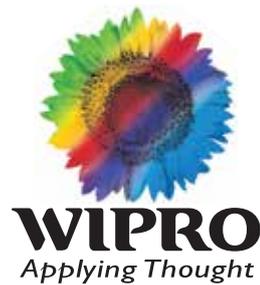
Conclusion

Despite the abundance of technologies and processes in the market today, there is a need for consolidation and industrialization throughout the entire Big Data lifecycle. Currently, there are no solutions available that provide a holistic approach to handle the various challenges faced by enterprises implementing Big Data. Moreover, there is clear need to identify a technology stack that will allow enterprises to aggregate, ingest, analyze and process, and consume data effectively.

But not all is hazy. There are solutions available that help organizations create value in their Big Data implementations. These solutions establish a measurable and repeatable methodology for various technical and process-related challenges and help identify the key activities that require maximum attention. Validation and verification of these activities will ensure that enterprises can extract value from their Big Data implementations that are also imperative for any solution. Enterprises need to identify the right solution that fits their Big Data implementation strategy. This will ensure that they find light in the chaos.

Survey Methodology

In May 2015, Wipro conducted a survey amongst senior executives and analysts based in USA & Canada (52%), Europe (33%) and Asia (12%) spread across various industries – Banking and Insurance, Retail, Consumer goods, Hi-Tech, Telecom and Utilities. Each of the enterprises interviewed had more than 1000 employees and over \$500M as annual turnover. The sample included CIOs responsible for strategic and financial plans (38%) and Vice-presidents responsible for implementation of technologies (48%) and Analysts who track the Big Data space on a regular basis (15%).



About Wipro Ltd.

Wipro Ltd. (NYSE:WIT) is a leading information technology, consulting and business process services company that delivers solutions to enable its clients do business better. Wipro delivers winning business outcomes through its deep industry experience and a 360 degree view of "Business through Technology." By combining digital strategy, customer centric design, advanced analytics and product engineering approach, Wipro helps its clients create successful and adaptive businesses. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, Wipro has a dedicated workforce of over 160,000, serving clients in 175+ cities across 6 continents.

For more information, please visit www.wipro.com

DO BUSINESS BETTER

CONSULTING | SYSTEM INTEGRATION | BUSINESS PROCESS SERVICES

WIPRO LIMITED, DODDAKANNELLI, SARJAPUR ROAD, BANGALORE - 560 035, INDIA. TEL : +91 (80) 2844 0011, FAX : +91(80) 2844 0256, Email: info@wipro.com

North America Canada Brazil Mexico Argentina United Kingdom Germany France Switzerland Nordic Region Poland Austria Benelux Portugal Romania Africa Middle East India China Japan Philippines Singapore Malaysia South Korea Australia New Zealand