Ensure Big Returns from Big Data with Effective Quality Assurance Strategy

Abstract

Data has always been an integral part of business. Technology transformation has further reinforced this importance. The advent of social media, mobility, and the Internet of Things has not only blurred the lines between online and offline business, but also resulted in the availability of a huge amount of unstructured data related to consumer behavior and interests. With big insights, Big Data could indeed mean big business for your company.

However, industry analysts are echoing growing concerns on data quality. Quality Assurance (QA) can offer a solution to the data quality challenge, and a robust and effective QA strategy can deliver big returns from Big Data.

The big deal is that with Big Data, your business can identify demand signals, track supply chain, and meet the right demand, at the right time, with the right price.

Big Challenges and Needs Analysis

Until now, Moore's Law on doubling of computing power every 18 months was considered a breakthrough. But with over a billion social media posts every two days, Big Data is changing definitions and benchmarks for breakthrough technologies. The need to stream and collect real time data from varied data sources and in different formats results in an exponential increase in volume, resembling a data tsunami. Further, there is also the need to continuously 'listen' to the stream and sift out irrelevant data. There is no point in responding to a customer tweet after a week, or having dirty, noisy data affecting business decisions.

The Mindset Change From Data Warehousing to Big Data

Traditionally, companies based their business strategy on structured data from enterprise applications and relational databases. Given Big Data's 'Variety' challenge, relational databases that store structured information are ill-suited to storing, analyzing, and processing Big Data.

While data-driven planning has been recognized as 'the best practice' in business, Big Data is changing some age-old notions of management. It is often confused with data warehousing.

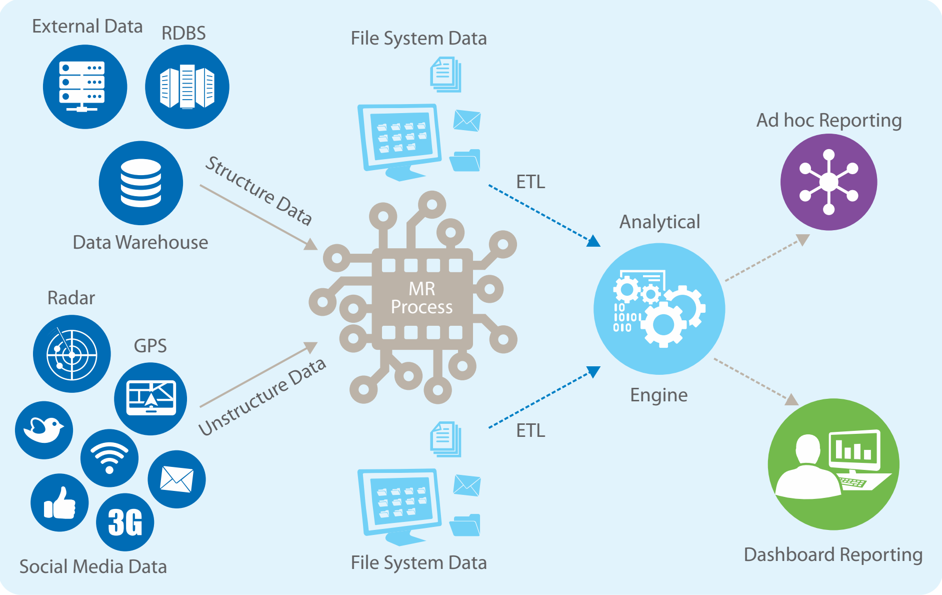
Parameter	Data Warehousing	Big Data
Size	Gigabytes	Petabytes or Zetabytes
Volume	Linear	Exponential
Format	Structured	Unconstrained
Application	Analyze Existing Business	Create New Business
Processing	Batch Mode	Real Time Streaming

Key differences between data warehousing and Big Data—while data warehousing provides answers to business problems, Big Data gives intelligent insights on the questions that matter.

From an assurance perspective, we too need to go the extra mile to assure Big Data quality. While we have well-defined, time-tested assurance and testing strategies for data warehousing, these traditional approaches, when directly applied to Big Data could result in compelling insights that are outright wrong!

Big Data testing requires test strategies for structured, semi-structured, and unstructured data.

Other pre-requisites include statistical tests on data sets, an optimal and scalable test environment, and the competency to work with non-relational databases. We need an efficient assurance framework for Big Data.



The Big Data Ecosystem—Big Data comes from various sources such as databases, enterprise systems, social media, global positioning systems (GPS), radars, radio frequency identification devices (RFIDs) or satellites. The data is processed through a multi-node MapReduce processing layer, which resides in a file system instead of a structured database. The processed data is then loaded onto an analytical engine, which transforms it to a readable format for business reporting.

The Four Pillar Framework

Big Data processing involves introducing new steps to the lifecycle that increase the complexity of the process. This complexity introduces four types of inaccuracies in the Big Data processing cycle:

1. Wrong data sources during requirements planning, resulting in data mismatch and wrong inputs for processing
2. Data inaccuracies resulting from the multi-mode MapReduce process that operates across file systems, requiring adequate configuration testing
3. Coding defects detected after comparing results of the transformation and business logic
4. Process and data gaps in the visualization layer

These inaccuracies can impact the quality of Big Data analytics, resulting in inaccurate insights and wrong business decisions. An efficient testing strategy, comprising adequate quality checks and controls is required to mitigate such risks. If left to ad hoc implementation, some of these controls may be missed. In addition, such implementations may appear complex and daunting. It is therefore important to manage these complexities and address the challenges with a robust quality assurance framework that is easy to implement, without compromising on the control rigor.

A framework based on four pillars that adequately supports quality assurance teams, addresses specific challenges, and reduces the complexity of Big Data processing to make it simple, not daunting:

1. **The First Assurance Pillar—People**

   Like any IT implementation, in Big Data too, quality is the responsibility of all stakeholders and not just QA teams. Effective implementation requires focus on data quality, which in turn, calls for additional quality assurance roles in the four phases of the Software Development Lifecycle (SDLC).

2. **The Second Assurance Pillar—Process**

   Big Data processing involves specific activities across four phases. The complexity of these activities introduces some inaccuracies in the process output. It is important to address these inaccuracies early in the cycle, through an effective assurance and testing strategy, simply because the cost of fixing defects in production is high. A process based approach helps identify inaccuracies to improve the overall effectiveness of the project.

3. **The Third Assurance Pillar—Infrastructure**

   The success of Big Data testing depends on executing the right test scenarios, which in turn, demands the right testing environment. Further, testing teams have limited budget and tight schedules. The right, scalable testing environment needs to be provisioned quickly within budgetary and schedule constraints. Cloud-based infrastructure plays a big role here. A combination of public, private, and hybrid cloud infrastructure can adequately address the computing, storage, software, security, and scalability requirements with zero or minimal upfront infrastructure cost—a big boon for testing teams with limited budget.

4. The Fourth Assurance Pillar—Automation

Automated testing is the best way to increase effectiveness, efficiency, and coverage in Big Data testing. Real-time listening demands that testing teams be agile. The fourth pillar proposes that automation tests be conducted in parallel with development. Automation scripts and tools for data profiling, data validation, test data generation, test data base-lining, predictive analysis, and modeling ensure the correctness of the implementation and drive significant cost savings.

Big Benefits

The four pillars, when working in unison, offer the potential to drive huge cost and effort savings, and also win user trust. The benefits include:

- Improved productivity
- Defect prevention and early flagging of defects
- Reduced cost of data quality
- Zero or minimal upfront investment
- Quick creation of testing environment
- Agility of testing teams
- Scalable storage and computing infrastructure
- Faster turnaround time

Conclusion

The recommended approach is innovative and effective to crunch large volumes of Big Data. It can help determine consumer buying patterns and increase sales with 'Also Buy' suggestions, analyze crime trends to predict and prevent crime, associate climate and crop health data for better crop cultivation, or even predict the outbreak of epidemics—through means that are faster than hospital admission records. To prevent disasters that could result from basing these decisions on inaccurate or flawed data, it is critical to leverage the framework and its four assurance pillars.

With effective QA, you can use Big Data insights to pre-empt competition, explore new markets, predict consumer behavior, drive marketing, and improve brand value—important factors to business' bottom line and agility.

About The Authors

Tom Edwards

Tom Edwards is Senior Vice President, Technology at Nielsen. With over 18 years in Software Development and Quality Assurance, Tom has vast exposure in Business Intelligence Testing methodologies.

Rajni Sachan

Rajni Sachan is Business Relationship Manager at TCS. She has spent 14 years in Business Intelligence and Quality Assurance, implementing strategic and complex IT transformation programs across domains.

About TCS' Assurance Services Unit

With one of the most comprehensive testing portfolios on offer, TCS addresses business, quality, and risk management challenges for its global clients. We empower organizations across domains to ensure first-time-right releases, reduce cost of quality, and accelerate superior customer experience.

TCS offers assurance services across the testing value cycle, including test consulting and advisory, test services implementation, and managed services for test environment and test data management. We continually redefine testing and QA paradigms to help our clients stay ahead of the curve. Our library of domain-based reusable business functions and proven engagement model founded on the twin pillars of product and process quality enable us to deliver certainty to our clients.

Over 30,000 testing consultants, strategic alliances and partnerships with key product vendors, more than 65 dedicated test centers of excellence and our innovation labs power our tailor-made solutions, testing assets and accelerators. TCS has the unique distinction of being the only global IT services organization to have been recognized consistently as a 'Leader' in the Quality Assurance space by leading analyst firms – Gartner, Ovum, Everest, IDC, NelsonHall and HfS - over the last few years. With specialized test environments and labs, TCS drives the delivery of assurance in a non-disruptive, agile, and automated manner, making the entire development lifecycle more efficient.

Contact

Visit TCS’ Assurance Services unit page for more information

Email: global.assurance@tcs.com

Blog: #ThinkAssurance

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India’s largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

Experience certainty.

IT Services
Business Solutions
Consulting

TCS Design Services I M I 11 I 16