

White
PAPER

**Overcoming QA
challenges in Big Data
Analytics Space**



The data is growing at a very high speed. According to IBM research, approximately 90% of the data has been created in the last two years. This large chunk of data, also called big data has made the role of quality professionals more challenging than ever. A large number of big data initiatives have failed due to data quality issues and lack of proper quality management strategy. In this whitepaper, Test Triangle offers its insight to handle big data testing challenges.

Introduction

Big data technology allows capture and analyses of large amount of data generated by numerous sources. It provides us with powerful insights, which is useful in predictive analyses and business decision-making.

Today, business leaders are attracted towards big data because of its capability of embedded analytics. But big data initiatives are not quite successful today and people have started to consider it as intricate technology. According to Gartner (2017), close to 85% of big data initiatives fail. Most of the companies find it challenging to embed the modern big data practices on the existing infrastructure. By and large, current organizations' culture does not support data-driven decision-making. It has fuelled the demand of big data QA and testing, which can continuously monitor and validate big data development.

Big Data Testing

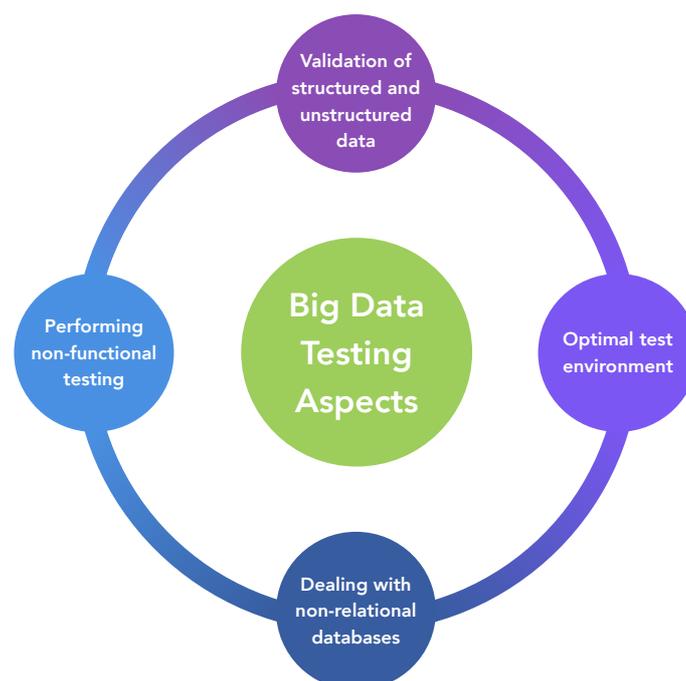
Big data is still in its nascent stage; however, market is flooded with tools that can bring higher efficiency and cost-effectiveness to the big data development. There are several popular tools such as Hadoop, HPCC, Storm and Qubole that work in the areas of extraction, storage, cleaning, mining and data visualization. These tools are very sophisticated and technically intricate; therefore, testing is important element of big data development.

There are several aspects to big data testing, specifically, validation, test-environment setup, non-functional testing and handling the non-relational database. If these testing aspects are not implemented properly, it leads to poor testing quality.

Therefore, test approach should also be defined early in the test lifecycle. In big data, performance and failover testing is necessary for evaluating the scalability of the test process. Big data resides in semi-structured or unstructured formats. Therefore, heterogeneity, volume and velocity of the data increase testing challenges. The major challenges in big data testing are:

- Heterogeneity and Incompleteness of data
- High Scalability
- Test Data Management

Moreover, the current software testing methodologies are adapted to test homogenous and clean data.



Handling Big Data Testing Issues

Addressing Performance Issues

Parallelism is the most common approach to handle the performance issues. The latest databases can achieve parallelism in two ways. In the first approach, the information is divided into intelligent segments. In this approach, different database operations are performed simultaneously. However, if the two operations are using the same information, they should work in a serial manner. This approach can be used to handle overwhelming workload. In the second approach, the application information is intelligently segmented. NoSQL handles all kind of data. It can be used to store unstructured data.

Indexing is another method to handle the performance issues. In indexing, the records are sorted on multiple fields. In this approach, another database is required for holding the pointer to the record. NoSQL can be used for creating this index.



Addressing the Scalability Issues

There are several techniques to handle the scalability issues in big data testing.

Clustering techniques: In the clustering technique, the large amount of data is distributed equally among all the nodes of a cluster. With this technique, large data files can be easily split into different chunks and stored in different nodes of a cluster. The file chunks are replicated and stored in different nodes to reduce the machine dependency. Hadoop is also a clustered database. Hadoop programs can be easily scaled up to a larger hardware with little changes in the program.

Data Partitioning: In the data partitioning, parallelism is conducted at the CPU level. It is less complex and offers easier execution.

Addressing the Test Data Management Challenges

In big data testing, there are several challenges in test data management such as managing the data in automation testing, acquisition and management of test data, and maintaining accuracy and heterogeneity in test data. There are several approaches for handling the big data challenges in testing environment:

Planning and designing:

In big data testing, the automated test scripts cannot be scaled up for testing large amount of data. If these scripts are scaled up to test big data, it will lead to delayed response time and timed-out execution. The action-based testing can be used to handle this challenge. In this testing, the test scripts are considered as actions. These actions are easily scalable.

Infrastructure Setup:

In big data, test automation consumes a large amount of resources. Instead of setting up a dedicated infrastructure, the companies can use cloud technology.

Addressing functional issues:

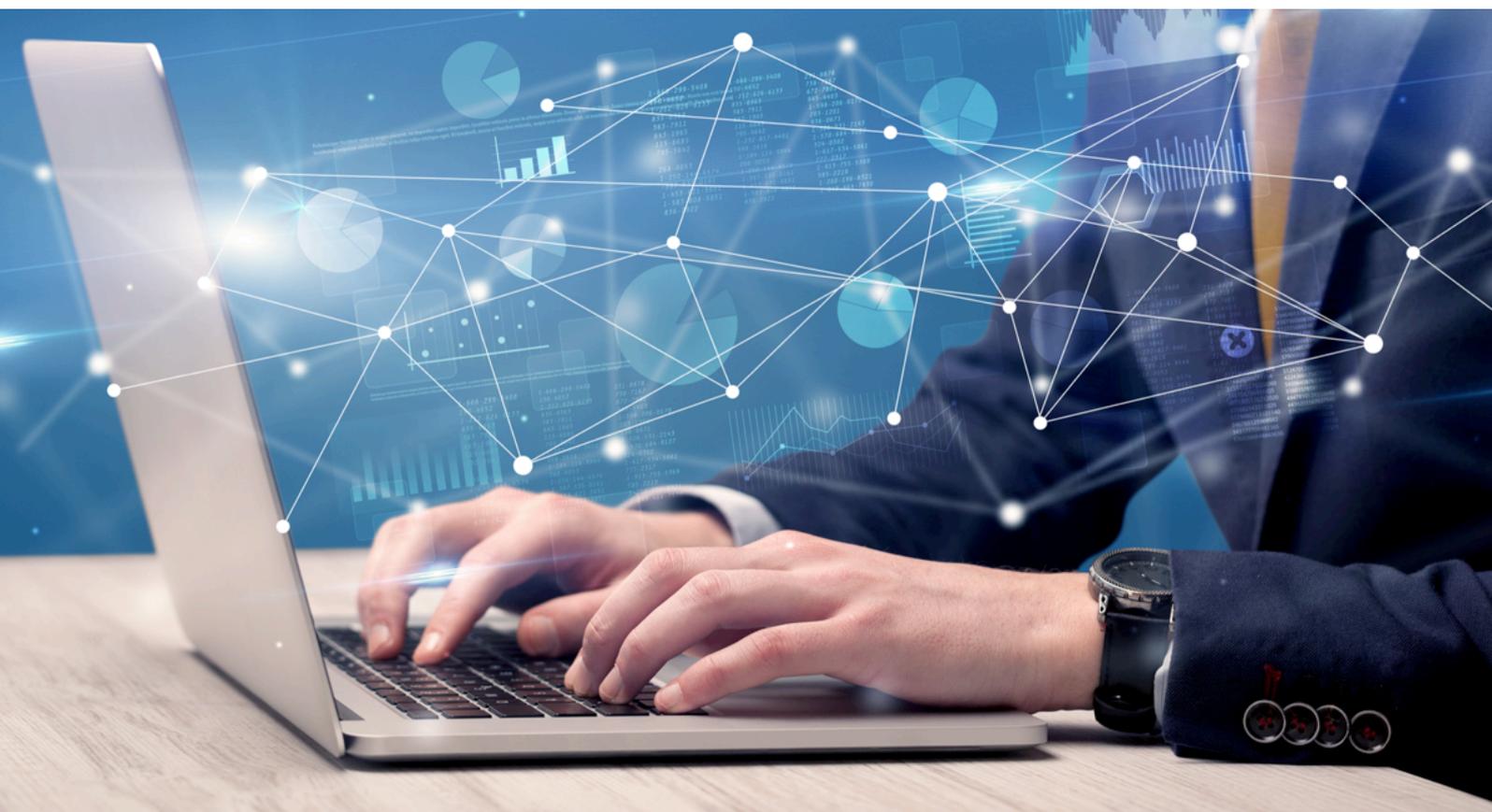
In big data, there are several functional and data accuracy issues arising from incorrect data replication, processing, and untimely transmission of results.

Data flow validation is appropriate solution, where it is checked that the data is acquired in accordance to the business cases. The data movement should also be validated in different layers. The data aggregation and data filtering mechanisms should be authenticated for proper functioning.

Future trends in Big Data Testing

The big data market is growing at an outrageously high number. The testing process in big data applications is also evolving at the same pace. Following upcoming trends are predicted in the future:

- In the future, there will be advanced strategies to generate the giant quantity of test data and reduce the big data test set-up time and cost. Specialized test tools and creative exploratory approaches will be used for data error trapping and business rule implementation.
- The data quality will be a critical issue in the near future. There will be various benchmarks for evaluating the data quality such as conformity, accuracy, consistency, validity, duplication and data completeness.
- Live data integration will be used to improve big data testing efficiency. It will be used to verify the data before it is moved in database.
- There will be novel strategies to gauge the performance of big data applications. In future, performance testing methodologies will integrate the statistical analysis from different application layers.



About Test Triangle

Originally founded in 2012, Test Triangle has become a leader in IT consultancy services providing services in application testing, DevOps, RPA, Custom software development, mobile app development, Atlassian consultancy, niche IT staff augmentation and training in advanced technologies. Test Triangle is headquartered in Ireland; but it also has branch offices in London, United Kingdom, and Hyderabad, India. We have exponentially grown to become a team of 200+ members providing services in different verticals such as Banking & Finance, Utilities, Pharma, Retail, IT & Education etc.

Test Triangle's R&D department has created a propriety platform, Test Outsourcing Dashboard [TOD] which can be used to manage software testing lifecycle using collaboration tools like email, live chat, video conferencing. We have also launched a self- service testing platform (the premium version will be released as SaaS solution), which can provide a project overview and real-time updates of the software development lifecycle.

Over the years, we have established the reputation of being a 'trusted partner in IT consulting'. Test triangle is an agile software company, which constantly strives to exceed the expectations of its clients. We adopt the software testing and software application lifecycle to meet the customer's demand in an efficient and reliable manner. With a global workforce, we have proved ourselves in delivering tight-deadline projects.

We are proud to declare ourselves a client of Enterprise Ireland and European commission.



For inquiry please contact: inquiry@testtriangle.com

Ireland - HQ

Suite 12, Plaza 212 Blanchardstown Corporate Park,
Ballycoolen, Dublin, D15 W535

UK

4th floor, 86-90 Paul Street, London, EC2A 4NE

India

1-98/9/3, Plot No.3, Flat No.102, Jaihind Enclave,
Madhapur, Hyderabad 500 081

**Sales
Phone
Number**

ROI Hotline

+353 1 9685077

UK Hotline

+44 (0) 2071933020

India Hotline

+44 (0) 2071933020
+91 40 49510533



facebook.com/TestTriangle



linkedin.com/company/test-triangle



twitter.com/testtriangle



youtube.com/user/TestTriangle