[Automation](https://www.newgenapps.com/blog/robotic-process-automation-in-enterprises) of Big Data is the most troublesome technology which is totally changing the dominion. The main element of Big Data lies in finding out the patterns which actually have some projecting values. It is then developed by automated processes that help in identification of some particular ‘data features’. These features then assist in building a predictive investigation of the database. From the [automation](https://www.newgenapps.com/blog/what-is-robotic-process-automation-uses-application-benefits) of Big Data Analysis, a great number of organizations have achieved benefits. According to the technology applied any amount of Big Data can be processed, analysed and understood within few weeks.

Recently, the [researchers from MIT](http://www.engineering.com/DesignerEdge/DesignerEdgeArticles/ArticleID/13471/Automation-and-Big-Data-Analytics.aspx) conducted the test of Big Data Analytics by removing the human factor from its processing. The prototype called Data Science Machine was implemented in several Data Science contests where this automation performed better than its human competitors. Its level of accuracy was up to 96% in this test.

Interestingly, while humans took months for decoding their prediction algorithms, the machines did the same in just few hours. This research aimed at automating the Big Data Analysis includes preparation of specified data and identification of problems that can be resolved through this analysis.

In the processing of Big Data various tools, techniques and frameworks are involved. Data creation, storage, retrieval, and analysis are related to big data which is outstanding in terms of volume, diversity, and rate. Rather than testing the individual features of the software product testing [Big Data application](https://www.newgenapps.com/blog/big-data-in-marketing-applications-uses-benefits) is more of a verification of its data processing. Performance and functional testing are the keys to Big Data testing.

It is ***concluded*** that, there is lots of scope in doing automation of big data in today’s world. Most of the organisations are putting manual effort to test and verify their data either it is Facebook, LinkedIn, Instagram, Amazon etc. all big companies relied on manual testing because no end to end automation framework is available in market which suits their needs.

The automation of Big Data is a huge step in the direction of improving Data Science in the imminent times. Implementing entire framework will results in huge cost and effort save.

The ***proposed*** automation tool, caters end-to-end automation of any big data, it focuses on 4’Vs of big data as described in earlier chapters. Big Data Testing plays a vital role in Big Data Systems. If Big Data systems not appropriately tested, then it will affect business, and it will also become tough to understand the error, cause of the failure and where it occurs. Due to which finding the solution for the problem also becomes difficult. If Big Data Testing performed correctly, then it will prevent the wastage of resources in the future.

Automation of big data testing is the key to survive all industries in longer run. Data Warehouse testing is very important because the business relies upon this data to make key decisions. Bad data and issues in Data Warehouse can lead to failures of the Data Warehouse projects because lack of trust in the data and consequently low usage.

Some of the big organisations have already started developing their in-house tools to handle all manual testing involved in big data, but that tool will be specific to that organisation and will fulfil their needs.

The ***proposed*** data warehouse automation tool, can handle all the testing involves in big data and it will be open source tool which can be used by all small and big organisation without paying any extra cost and can save huge amount of manual effort by automating their tasks. It is built specifically to automate the testing of Data Warehouses & Big Data. It can automate the entire process for you and fit nicely into your DevOps strategy.

To ***summarize*** the research work, in brief, an automation framework has been designed and developed to handle data warehouse automation testing. A POC has been completed at all levels of testing and this framework gives greater results as compare to existing tools and even coverage is huge as it covers performance and security testing as well.

This automation tool has been designed and developed in RBS labs and successfully implemented in all phases of data warehouse testing. Data warehouse automation testing at all levels either it is functional or non-functional is being done by this tool after taking some sample of data. So in summary, if any organisation big or small implements this framework in their area, they will get huge benefit out of it as they can cut down their major manual efforts and can save huge cost. Even performance and security testing are big achievements which can be done automatically via this tool.

***Future*** scope of this research work will make this framework more robust. This research work is more focused on current data warehouse testing issues which every organisation is facing on day to day basis. This research is mainly focused on structured data. In future, this framework can be further extended to big data tools like Hadoop, TeraData and NoSQL to handle un-structured data as well.

The proposed work can be improved in future in the following directions:

* Automation testing of un-structured data along with structured data. There are no automation tools or frameworks available to test un-structured data yet and most of the organisations have already started using these big data tools to handle their un-structured data, so there is a huge scope in extending this framework for automation testing of un-structured data as well.
* End user reporting is a major component of any business intelligence project. The report code may execute aggregate SQL queries against the data stored in data marts and/or the operational DW tables then display results in the required format. There can be multiple ways to automate these reports testing.
* In this work, automation is based on OLTP (Online transaction processing) design framework, in future, this can be extended to OLAP (Online analytical processing), where we can make data source of dimensions and cubes.
* In this work, all reports are in tabular format. In future work, reports can be more customised and can be at more granular level where users can see each and every detail. Even users should be able to schedule their test automation at particular time for their regression testing.