



Data and Analytics Testing on BI Platforms

A Path to Successful Model-Based Testing Approach



Analytics and Business Intelligence (BI) platforms facilitate the strategy of combining data integration and analytics technologies to create data analytics solutions that enable informed and actionable business decisions. The combination of Big Data, fast data, modelled data, and predictive analytics on modern BI platforms not only allows for fast analytics of historical and current information, but also provides the capability to predict future scenarios. Many organizations continue to adopt new technologies in pursuit of achieving further capability to capture, transform, and visualize vast volumes of structured and unstructured data in a timely manner. The ability to expedite data into business information provides organizations with capability to quickly capitalize on the knowledge, utilizing insight to enable strategies which may result in, but not limited to, a competitive advantage, more productive operations, or predictions of customer behaviors.

What is Data and Analytics Testing on BI Platforms

In the age of big and fast data, the quality of data is paramount to the success of any data integration and analytics implementation. With an abundance of the types and volume of data, it is more crucial than ever to ensure that the correctness of data is not compromised in the endeavor to achieve fast analytical decisions. It is imperative that testing on BI Platforms is performed at various stages of the data integration, transformation, and presentation to ensure that information, used to provide insight and make business decisions, is accurate, complete, and timely.

Good testing within BI platforms identifies issues early, prevents data loss or inaccuracies, and validates the quality of the solution. Quality assurance activities should be performed at the following key stages of data processing:

- Source data validation i.e. source data files, logs, and media
- Data ingestion of structured and unstructured data into a landing and/or staging area
- The various transformation processes of turning raw data into predictive analytics and structured business data i.e. ETL (Extract, Transform, Load)
- Data visualization and presentation layer i.e. Power BI, Qlik

Within the stages of testing, the following areas, where applicable, should be included for validation:

Data Definitions and Profiling

- Metadata validation of nulls, data formats, field size and range, and precision constraints.
- Schema validation such as consistent structure and use of business keys and unique indexes to enforce record uniqueness.

Data Reconciliation

- Comparison of the number files and records loaded or processed between source and target system and tables to ensure completeness.
- Row-by-row comparison to identify record and attribute exceptions between source and target tables to identify processing and transformation anomalies.

Data Presentation

- Display, search, and filtering parameters and options for the reports work as intended.
- The information on the report is formatted correctly.
- The information is displayed as per the design specification and user requirements.
- The information reconciles with expected values in the underlying source(s).

Data Integrity

- Data models adhere to design structures and concepts
 - > Correct versioning or changed row capture of slowly changing dimensions, late arriving data, ODS (Operational Data Stores) or Data Vault Hubs to ensure conformance of data modelling concepts.
 - > Prevention of orphaned records in relational tables i.e. fact tables, to maintain correct and consistent referential integrity.
 - > Correct lookup, translation, and use of master data and reference data.
- Validate data is not lost through poor ETL job dependency and design.

Traditional Approach for Data and Analytics Testing on BI Platforms

Comprehensive testing of data and analytics can be a costly and time-consuming exercise, and oftentimes seen as disproportionately expensive when compared to the development effort. To effectively test data processing and transformation activities, good industry standards emphasize that it is best practice to test, identify, and manage issues and errors early in the data integration or ETL process to reduce complexity and effort required to resolve.

To achieve this, testing in the data and analytics domain has generally required either manual or scripted efforts and technicians, skilled in the use of query languages i.e. SQL (Structured Query Language) to craft and maintain scripts, domain knowledge of ETL and data modelling concepts, and thorough understanding of complicated business rules that have been applied within data transformations. Manual or script-based testing can be cumbersome, add maintenance overhead, and reduces the speed to delivery.

Model-Based Data and Analytics Testing on BI Platforms

Model-based testing within BI Platforms is a testing technique that uses a model to predict expected behaviors and outcomes of the data transformation process. A model-based approach removes or reduces manual and script-based test activities and instead, focuses on identifying the key states and expected results of data as it moves through the processing pipeline. In this way, the model can be used to predefine and automate expected results of an ETL process, including but not limited to, data values or lists created from a set of data transformations, the data types and formats of the attributes of input or output files, the expected referential integrity constraints of a data model, and the volumes of data or files expected during processing. A model-based testing approach on BI Platforms can provide the following advantages:

Increases the Speed of Testing and Speed to Delivery:

Reduces or eliminates manual or script-orientated testing activities, therefore expediting the test process. This results in a more comprehensive test suite that helps to build business confidence for the data, while simultaneously allowing for faster, incremental validation and adoption of analytics and BI solutions.

Better Use of Time, Focus on High Value Work:

Allows testers and testing teams to prioritize testing strategy, framework design, create intelligent and efficient testing approaches over the need to have specialized data integration, analytics, and BI domain knowledge and scripting skills. More emphasize can be placed on analysis and testing of critical or vulnerable areas such as data security and non-compliant data.

Greater Test Coverage and Volume:

Shifting from a complicated and time-consuming manual or script-based approach to a model-based approach allows for greater test coverage and quality. Testing can readily be performed on vast numbers of objects and at various stages of the analytics and BI solution without compromise to speed.

Simplifies the Testing Approach:

Model-based test automation tools employ more intuitive interfaces and reduces programming efforts, resulting in greater comprehension of the testing activity, and consequently enabling more cross-functional teams.

Better Monitoring and Understanding of Change-Impact:

Simplified critical tests which form part of a comprehensive automated test regression suite can provide better understanding of downstream change-impact resulting from design modifications of data models, addition or removal of source systems, or any changes upstream of the test validation point.

Industry Trend Driving a Shift Towards the Need for Faster, More Comprehensive Data and Analytics Testing

Information Security and Regulatory Compliance:

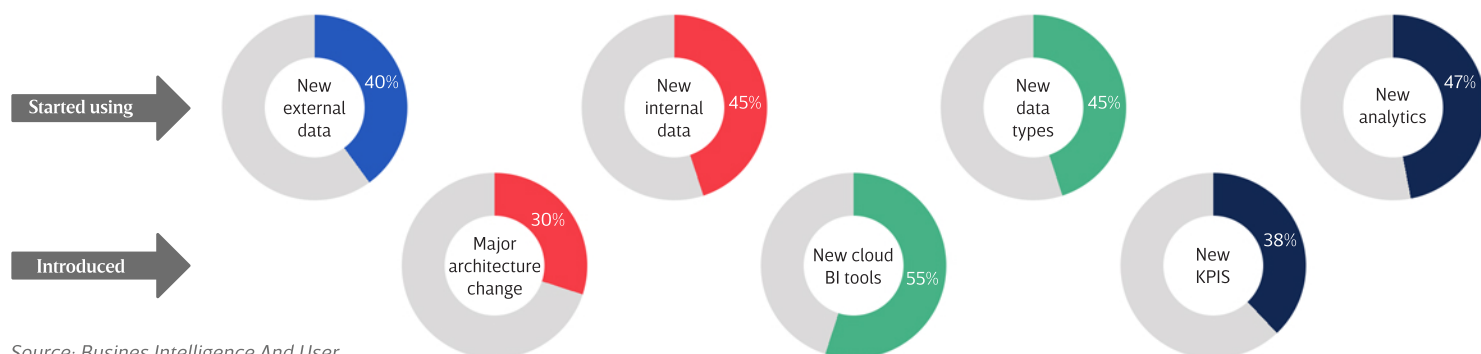
Data privacy and security regulations like the European GDPR (General Data Protection Regulation) law necessitates organizations to better understand data lineage and where there may be security breaches or exposure of sensitive data. By nature, analytics and BI solutions are susceptible to security risks because sensitive information is likely captured for analytics. Testing, identifying, and resolving potential exposure of high-risk data elements decreases an organizations' legal and business liability risks.

Continuous Integration and Smaller, More Frequent Releases:

Continuous delivery and deployment of smaller analytics and BI components based on DevOps and DataOps principals, requires continuous integration and quality assurance. The velocity and iterative aspect of continuous delivery dictates that this is only achievable if testing can be developed rapidly, is automated, and easily reusable and repeatable.

What Data and Analytics Changes are Affecting Enterprises?

In the last 12-18 months has your organization done any of the following?



Source: *Business Intelligence And User Survey*, IDC, February 2020, N=310

Path to a Successful Model-Based Testing Approach for BI Platforms

Test Maturity

A fundamental step to better testing is to identify and acknowledge test related issues and challenges in the data and analytics domain. Taking the initial step to evaluate the test maturity of the BI Platform test processes helps to identify areas that require focus and improvements. Consequently, the evaluation will determine test readiness to utilize model-based testing techniques. Using industry standard approaches like the Test Process Improvement (TPI) Next model, will help to improve test processes and progress through the following levels:

Initial: Ad hoc testing activities with little visibility into the risk and quality of the test deliverables and a high dependency on key individual(s).

At this test maturity level, all effort should be focused on improving test processes prior to adoption of any advanced testing techniques. Laying a strong foundation of good test practices paves the way for easier adoption and integration of new tools or techniques.

Controlled: Doing the right things. Test activities are done in a planned manner and there is better insight into the risks and the quality of the deliverable.

Some effort is required to further advance test processes. However, adoption of a model-based testing approach on a smaller scale, in areas with highly repetitive and uncomplex data processing activities like loading Operational Data Stores (ODS) will improve test resource efficiency.

Efficient: Doing things the right way. Once the right activities are being done, the next step is to improve how they are done.

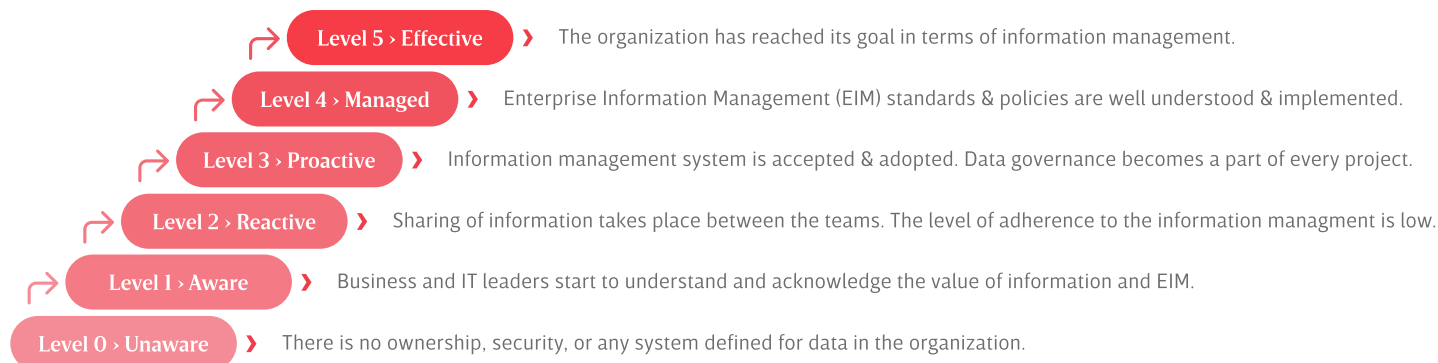
Ready to fully adopt a large-scale model-based testing approach to further enhance existing test capability. Test models can be implemented to cover the full data processing and development pipeline and improve test efficiency, coverage, and speed to delivery.

Optimizing: Continuously adapting to ever-changing circumstances with focus shifting towards ensuring that this level and quality of testing is maintained.

At this maturity level, adoption of a full-scale, model-based approach not only provides the benefits of efficiency, coverage, and speed, but also enables the formation of more complete test regression suites with complex test models that predict negative test paths and change-impact scenarios.

Data Governance

In conjunction with test maturity and to fully realize the benefits and successfully implement a model-based test approach on BI Platforms, a clear data governance framework is required. Data governance defines the people, groups, and technologies impacted or required to achieve the corporate data strategy. Within the framework, guidance on data risk and security, access controls, data processes and methodologies are outlined. A proactive corporate data governance level or higher denotes a good data strategy which can pave the way for successful end-to-end BI Platform testing strategies.



Tool Selection

It is important to evaluate and select a model-based test tool which meets the current BI Platform requirements and scenarios, but which is also capable of accommodating future technology changes or growth strategies. The evaluation will help to identify the right tool for the purpose and future-proof the investment. The following key criteria and questions should be considered during the selection process:

Security

- Does it align with the data security and data management policies outlined for the organization?
- What data and environment, and associated security constraints need to be considered?

Scalability and Flexibility

- Will it function for the current set of requirements and scale as these requirements grow or evolve?
- Is it possible, and what effort is required to adapt existing capability to the ever-changing technology landscape?
- How progressive is the tool and how innovative is the product compared to other similar offerings? Will it have longevity in the market and continue to be useful in the future?

Integration

- How easy is it to integrate with the existing set of tools and technologies?
- How easy is it to deploy and propagate to the required teams in the organization?

Cost and Support

- Do the implementation costs, maintenance fees, licensing and payment structures fall within budgetary requirements for adoption and sustained usage?
- What support is offered and at what level, how good is it, and what will it cost?



Moving from traditional data and analytics testing methods to a model-based technique on BI Platforms requires a considered approach. Effort is needed to assess an organization and test team's readiness to adopt a model-based test approach. However, once done, it enables ease of implementation and ultimately provides better returns on investment, while also empowering test teams as they engage in more thoughtful, higher-value test activities. Combining intelligent model-based testing with intelligent data and analytics solutions enables organizations to truly harness the power from having correct data at the right time.

Let's Talk

For more information, talk to a member of our team today.