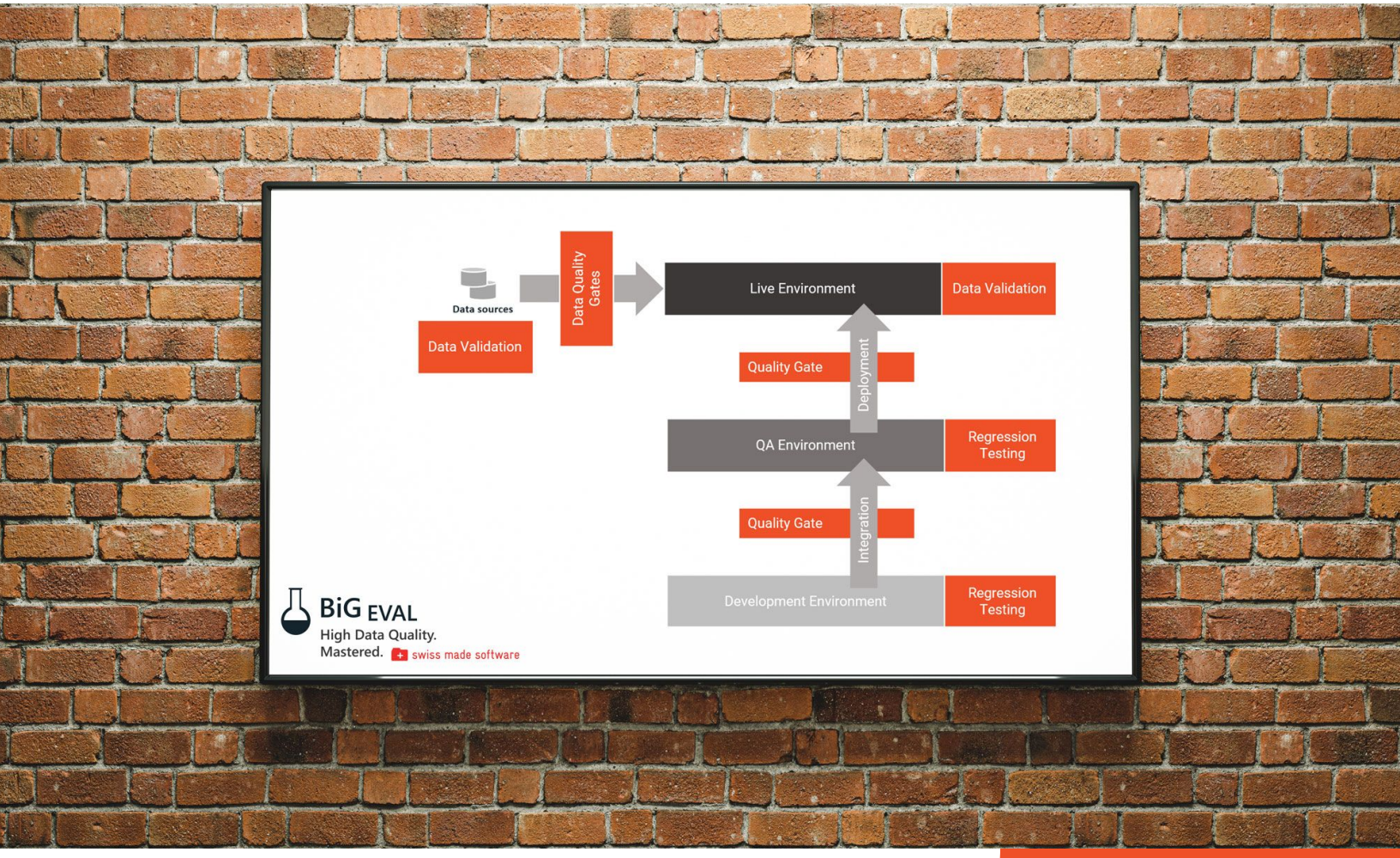
BiG EVAL
High Data Quality.
Mastered.

THE ULTIMATE GUIDE

# Creating a Winning Data Warehouse Test Concept

Thomas Bolt



BiG EVAL
High Data Quality.
Mastered.

# Table of Contents


4 Introduction to Data Warehouse Testing

7 Common Challenges

9 The Simplified Approach

10 Basic Test Cases

13 Advanced Test Cases

16 Test Automation

# What is this eBook about?

Data warehouses play a critical role in modern organizations, serving as the central repository for large amounts of structured data used to support decision-making and strategic planning. However, the accuracy and reliability of this data is only as good as the testing that has been performed on the data warehouse. This is why data warehouse testing is an essential aspect of data management.

This eBook outlines a simple approach to creating a robust and reliable data warehouse test concept, what it should cover, and how to gain efficiency and convenience with data test automation.

# Introduction to Data Warehouse Testing

In this chapter, we will introduce the concept of data warehouse testing and explore its importance in ensuring the accuracy and reliability of data. We will discuss the common challenges associated with testing data warehouses, including the complexity of the data, and how these challenges can be overcome. Additionally, we will examine the benefits of data warehouse testing, including improved data accuracy, increased confidence in decision-making, and reduced risk of errors.
One key factor that can make data warehouse testing easier and more efficient is the use of a data testing automation software. These tools can help to build simple test cases with minimal effort, freeing up valuable time and resources for other aspects of the testing process. That's why we will have a look at different technologies.
Finally, we will introduce the simplified approach to data warehouse testing that is the focus of this ebook, emphasizing the importance of starting with basic test cases and gradually adding more complex tests.

BiG EVAL
High Data Quality.
Mastered.

# Goals of Data Warehouse Testing

Data warehouse testing is the process of evaluating the accuracy and reliability of a data warehouse and its system components. The goals of data warehouse testing include:

**Ensuring data accuracy** Verify that the data stored in the data warehouse is accurate, consistent, and reliable.

**Validating business logic** Test that calculations, aggregations, and other business rules are applied correctly and consistently across all data sets.

**Testing performance and scalability** Ensure that the data warehouse can perform optimally and scale to meet the needs of the organization.

**Ensuring regulatory compliance** Test data privacy and security controls to ensure that the data is protected and meets legal and regulatory obligations.

BiG EVAL
High Data Quality. Mastered.

# The Importance of Data Warehouse Testing

Inaccurate data can have serious consequences for organizations, leading to a range of problems that can impact decision-making, operations, and reputation. The most common problems companies face include:

## Poor Decision Making

Inaccurate data can lead to incorrect decisions and wasted resources, as decision-makers rely on incorrect information to make critical choices. This can result in ineffective strategies, and a general lack of confidence in the data used to make decisions.

## Lost Opportunities

Incorrect data can lead to missed opportunities or ineffective strategies, as organizations make decisions based on false information. This can result in lost revenue, reduced efficiency, and wasted resources.

## Reputational Damage

When inaccurate data is released to the public or stakeholders, it can damage an organization's reputation and credibility. This can lead to reduced trust in the organization and its data, as well as negative consequences for the brand and its image.

## Inefficient Operations

Inaccurate data can also lead to inefficient operations, as processes are based on incorrect information and resources are misallocated. This can result in increased costs, reduced productivity, and a general lack of efficiency in the organization.

## Increased Risks

Inaccurate data can increase risk for an organization, particularly in industries with strict regulations and compliance requirements. This can result in legal or financial penalties, as well as damage to an organization's credibility.

# Common Challenges

Testing a data warehouse can be a complex and challenging process, due to the size and complexity of the data involved as well as the complexity of integration and transformation processes. Some common challenges that organizations face when testing their data warehouses include:

| | |
|---|---|
| Data Complexity | The sheer volume and complexity of the data stored in a data warehouse can make testing a daunting task. |
| Data Integration | Testing data warehouses can also involve testing the integration of data from multiple sources, which can be a complex and time-consuming process. |
| Data Quality | Errors and inaccuracies can be difficult to detect. This can include testing for missing data, duplicates, and outliers. |
| Data Security | Security breaches can result in the loss or theft of sensitive information. Testing includes checks for unauthorized access, data encryption, and data backup and recovery. |
| Performance | Slow performance can impact the ability of decision-makers to access and use the data. This can include testing query performance and data load performance. |
| Scalability | As data warehouses grow in size and complexity, it is important to test the scalability of the data warehouse to ensure it can handle increasing amounts of data. |

Data warehouse testing is a complex and challenging process that requires careful planning and execution. Understanding the common challenges associated with data warehouse testing is an important first step in developing an effective testing strategy. However, this challenge can be simplified and made easier by using data test automation software, which can streamline and automate many of the testing processes involved.

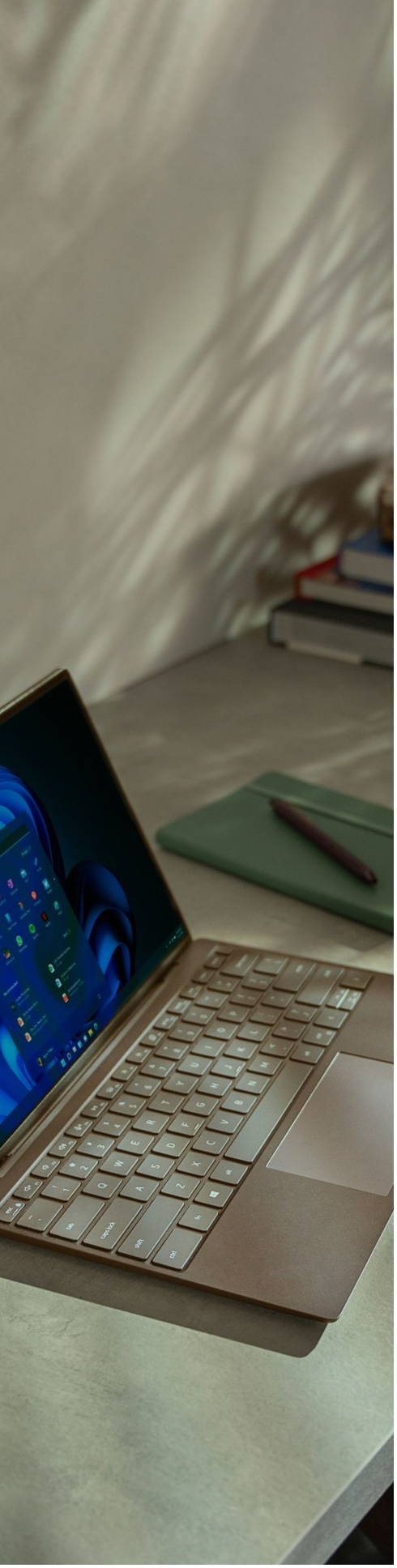
# The Simplified Approach

The focus of this ebook is a simplified approach to data warehouse testing that starts with basic test cases and gradually adds more complex tests. This approach has a number of benefits, including:

**Quick Results** Starting with basic test cases allows organizations to quickly identify and address any issues with the data warehouse, and see the benefits of testing without a lot of effort.

**Gradual Improvement** By gradually adding more complex tests, organizations can build a comprehensive testing strategy that covers all aspects of the data warehouse, and continuously improve the accuracy and reliability of the data.

**Step-by-Step Approach** Taking a step-by-step approach to testing helps organizations to avoid becoming overwhelmed by the complexity of the process, and ensures that they can focus on one aspect of the data warehouse at a time, reducing the risk of errors and increasing the accuracy of the testing process.



BiG EVAL
High Data Quality.
Mastered.

# Basic Test Cases

Using just some basic test cases are an essential starting point for data warehouse testing, providing organizations with quick and beneficial results with minimal effort. The following test categories, are used to validate the basic functionality of the data warehouse:

1. **Data Completeness** Verifying that all required data is present in the data warehouse, and that the data is complete and accurate.

2. **Data Consistency** Testing the consistency of the data, including verifying that the data is consistent across different sources, and that the relationships between data elements are accurate.

These basic test categories are an essential starting point for data warehouse testing, and build a solid foundation for more complex testing later on.

# Basic Test Cases for Data Completeness

Ensuring that the data stored in the data warehouse is complete and accurate is essential. Below are some of the most important basic test cases for checking for data completeness, which provide a good starting point for further validation later on.

- [ ] **Missing Data** Test to ensure that the number of records in the data warehouse matches the expected number of records and that all business keys are present. This can be done by comparing the number of records in the data warehouse to the number of records in the source system or to a previously established baseline.

- [ ] **Duplicated Data** Verifying that there is no duplicated data in the data warehouse, and that each record is unique.

- [ ] **Data Availability** Verifying that the data is available both technically and on time, including testing for data that is inaccessible due to technical issues or that is not updated in a timely manner.

- [ ] **Aggregation Validation** Test to ensure that data aggregation and summarization processes are working correctly and that no data is missing from the aggregation. This can be done by comparing the results of the aggregation process to the expected results and verifying that the expected level of detail is present in the aggregated data.

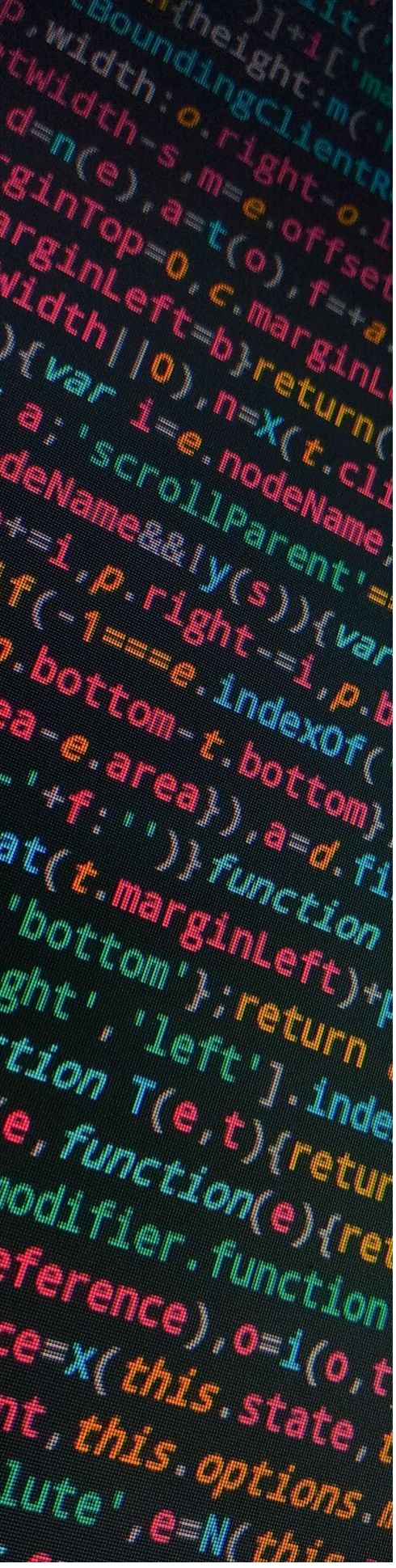
# Basic Test Cases for Data Consistency

Data consistency test cases ensure that the data is consistent across different sources and that the relationships between data elements are correct. This helps to ensure that the data is accurate and can be used effectively without any weird results in reporting.

- [ ] **Data Type Validation** Test to ensure that the data type of the fields in the data warehouse is consistent with the expected data types for those fields. For example, if a field is expected to store a date, the test case would ensure that only dates are stored in that field. This is even more important if any generic data types were used. E.g. a character-typed field contains numeric or date information.

- [ ] **Cross-Table Consistency** Test to ensure that data stored in different tables in the data warehouse that are related to each other are consistent. For example, if there is a customer table and an order table, a test case would ensure that a customer cannot have an order with an invalid or non-existent customer ID. This is especially important if the data storage technology doesn't provide referential integrity features or if they were disabled for any reason.

- [ ] **Historical Data Consistency** Test to ensure that data stored in the data warehouse over time is consistent. For example, if a data warehouse stores financial data for multiple years, a test case would ensure that calculations and aggregations based on historical data are correct and consistent across all years.



BiG EVAL
High Data Quality.
Mastered.

# Advanced Test Cases

As the data warehouse evolves and the testing process becomes more sophisticated, additional test cases can be added to provide a more comprehensive view of the data warehouse. We have grouped these test cases into the following categories:

1. **Technical Test Cases** focus on ensuring the technical reliability and functionality of the data warehouse

2. **Business Test Cases** focus on ensuring that the data warehouse meets the needs and requirements of the organization and its stakeholders.

# Advanced Technical Test Cases

As the data warehouse becomes more complex, additional technical test cases may be needed to identify and address potential issues, such as performance bottlenecks, integration problems, and data corruption, which can impact the reliability and usability of the data.

- [ ] **Performance testing** Evaluating the performance of the data warehouse under various conditions to identify bottlenecks and scalability issues.

- [ ] **Integration testing** Testing the data warehouse with other systems and applications to ensure that it is properly integrated into the organization's technology ecosystem.

- [ ] **Regression testing** Re-testing the data warehouse after changes or updates to ensure that existing functionality has not been affected. All test cases built in previous iterations - including the basic test cases mentioned before - are subject to be run over and over again.

- [ ] **Disaster recovery testing** Testing the ability of the data warehouse to recover from disasters, such as system failures or data corruption. Shutdown, restart the data warehouse system. Test partial and full data reload processes to ensure they work when needed.

# Advanced Business Test Cases

By testing the business aspects of the data warehouse, organizations can ensure that the data is accurate, complete, and secure, meets the needs of end-users, and is compliant with relevant regulations and standards. This can help ensure that the data warehouse is a reliable source of information for decision-making and strategic planning.

- [ ] **Data quality testing** Ensuring that the data stored in the data warehouse is accurate, consistent, and complete by testing for duplicate records, missing data, and inconsistent data.

- [ ] **Security testing** Testing for unauthorized access, data breaches, and other security vulnerabilities to ensure that the data warehouse is secure. Use a testing tool to impersonate another security context to automatically verify that users can only access data for which they are authorized.

- [ ] **User acceptance testing** Testing the data warehouse with end-users to ensure that it meets their needs and requirements. Make notes about how they use the tools. Ask them how and why they can decide whether a piece of information is right or wrong. Their completely different perspective and knowledge of the information presented is invaluable.

- [ ] **Compliance testing** Testing to ensure that the data warehouse is compliant with relevant regulations and standards, such as data privacy and security regulations. These tests also lend themselves very well to automation, as they enable you to confirm at any time that regulatory requirements are being met and even verified.

BiG EVAL
High Data Quality.
Mastered.

# Test Automation

Test automation is a powerful tool for improving the efficiency and effectiveness of data warehouse testing. Automated tests can help reduce the time and effort required to perform repetitive testing tasks, and can help identify issues more quickly and accurately. By automating tests, organizations can increase the speed and coverage of testing, while also reducing the risk of human error.

The benefits of test automation include:

**Increased testing speed and coverage** Automated tests can be run much more quickly and frequently than manual tests, increasing the speed and coverage of testing.

**Reduced testing effort and costs** Automated tests can save time and effort by eliminating the need for manual testing, reducing the overall costs of testing.

**Improved accuracy and reliability** Automated tests are less prone to human error than manual tests, improving the accuracy and reliability of test results.

**Faster identification of issues** Automated tests can help identify issues more quickly, allowing organizations to address them before they become critical.

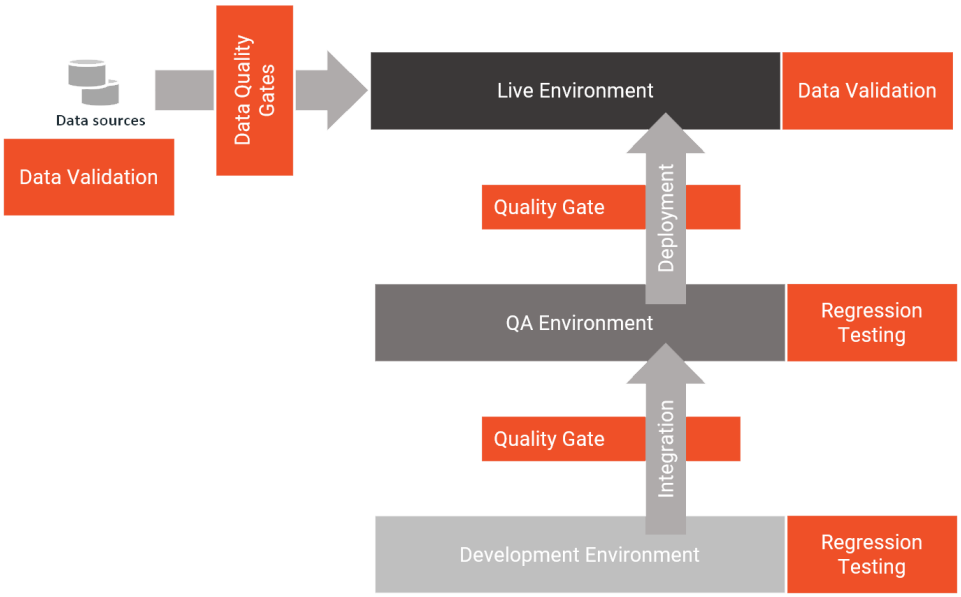
**Better test documentation and reporting** Automated tests can provide more detailed documentation and reporting, making it easier to track and manage test results.

BiG EVAL
High Data Quality.
Mastered.

# Test Automation with BiG EVAL

One tool available for automating data warehouse tests is BiG EVAL. BiG EVAL is a comprehensive data warehouse testing and validation tool that supports a wide range of data sources and test scenarios. With BiG EVAL, organizations can automate the testing of data consistency, data completeness, performance, and security, among other aspects of data warehouse testing.

BiG EVAL is also meant to be used in live data environments for ongoing data quality management during the whole life cycle of information.




## Want to Learn More About BiG EVAL?

Contact us today to learn more about how BiG EVAL can help you automate and optimize your data warehouse testing process.

info@bigeval.com

BiG EVAL
High Data Quality.
Mastered.

# Legal Information

**Author**
Thomas Bolt
Founder, CTO of BiG EVAL

**Publisher**
BiG EVAL, Inc.
70 W Madison St., Ste 5750
Chicago IL, 60602
United States

Web: https://bigeval.com
E-Mail: info@bigeval.com

**All rights reserved.**
This eBook may not be reproduced in whole or in part without the permission and precise citation of the author and publisher. All rights belong to the publisher or any named creators.