

Compliance to methodology and best practices in ETL solutions

Standardization quickly becomes an issue in heterogeneous environments with more than two or three ETL developers. The development guidelines and methodologies have to be set in order to keep the ETL solutions maintainable and extendable even in the distant future. These methodologies may contain hundreds of rules ranging from simple naming conventions up to very complex design patterns for different types of solutions.

Importance of code review

Before the developed ETL solutions are deployed, they should be reviewed thoroughly to ensure that all the guidelines have been honored. Otherwise, various problems can occur – ignoring a naming convention may make the code hard to maintain and forgetting to set an appropriate relational connection in a session will cause the solution to fail and may result in an SLA breach. Also Incorrect data types can result in data precision loss or data quality problems, such as foreign names becoming garbled. In our experience, it may be difficult to even detect a problem, let alone to find out where the damaged data has been propagated in the whole data warehouse environment (unless you have an advanced data flow analysis tool). So, code review is an important step in development process that can help in discovering and fixing the issues early and helps to increase the trust in the data integration process.

Benefits of Automation

Unfortunately, performing such checks manually is time consuming, exhausting and prone to errors because we are not fond of repetitious, tedious tasks - especially when a lot of rules are involved. For example, let us consider that verifying whether one solution adheres to one single rule takes 10 seconds on the average, and that we have 600 rules and 600 workflows¹. Under these circumstances, a complete manual check of the entire solution may take 6 months. Also, estimating the impact of one new rule in all workflows would be very slow and time consuming

The natural solution is to automate the process completely. This way, it becomes easy to perform complex checks of to-be-deployed ETL solutions in a matter of minutes or seconds, to avoid making mistakes or overlooking something, and to save a lot of time along the way.

Also, the development department can see the impact of a new rule almost immediately, and based on the findings of how many workflows or mappings do not conform to the new rule, they can estimate how time consuming and costly the repair would be.

Manta Checker - Code Review Automation Solution

The ETL development guidelines themselves will do no good unless they are enforced by an authority capable of performing fast, reliable and repeatable checks. Specialized tools and extensions to help you with this task are available in the Informatica Marketplace. One such tool is Manta Checker.

Manta Checker (formerly Profinit PSAT Checker) automates code reviews of Informatica PowerCenter workflows and mappings. Among its many unique features, Manta Checker can work disconnected from

¹ The estimate is based on an actual experience with a customer running a data warehouse containing 600 Informatica workflows.

the actual Informatica environment, thus allowing individual developers to run self-assessments on their own machines as frequently as needed before submitting their work for a formal evaluation.

Article by Erik Kratochvil, New Frontier Group, an Informatica partner and the producer of Manta Checker

Visit our Block on [Informatica Marketplace](#)

- [Manta Checker, Informatica PowerCenter Edition](#)