

WHAT IS PROACTIVE DATA QUALITY?

Proactive data quality is data quality that takes place before the data is put to use.

WHAT DOES THAT MEAN?

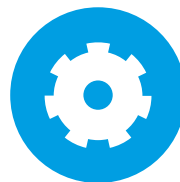
Data is in use when it is employed for



ANALYTICS



REPORTING



OPERATIONS

During these activities, people in the business are:

- Using the data to analyze the business' past
- Using the data to direct the business' future
- Using the data to interact with customers
- Trusting the data to be accurate

WHO DOES IT?

Proactive data quality is a partnership between:



**AUTOMATED
SOFTWARE-DRIVEN
OPERATIONS**



**EXPERT
HUMAN
INTERVENTION**



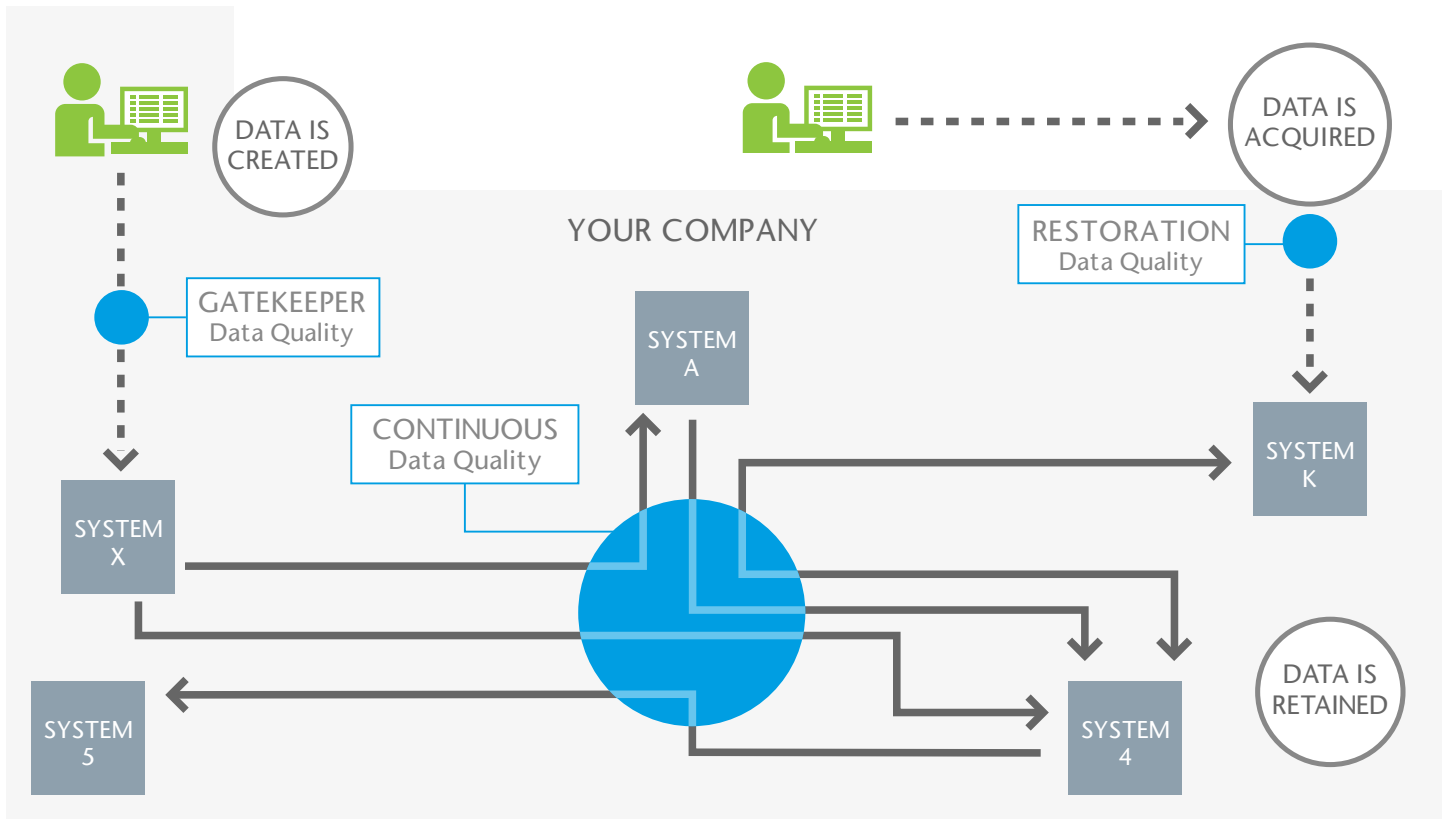
Automation does most of the work volume.

- Essential for keeping up with performance and timeliness demands with modern data volumes.
- Data quality requires sophisticated decision-making.
- The software must be able to represent complex processes so it can recreate detailed reasonings.

Experts resolve the issues that require a human touch.

- Human intervention is reserved for the small set of issues that can't be solved any other way.
- Some issues need external research that software isn't equipped to perform.
- Other types of fixes may need human verification or business judgment software can't exercise.

Where is it located?



Where data is CREATED (Gatekeeper)

Gatekeeper data quality prevents data from being created in a system if it does not pass certain checks. It tends to focus on validation and standardization of single fields.

Although it is relatively easy to implement gatekeeper data quality, gatekeeper data quality is inherently limited and can have negative operational effects. It also requires access to the system where data is created, which is not always available.

Where data is ACQUIRED (Restoration)

Restoration data quality performs data quality on data that has already been created, at the point where that data enters the company's control. It offers more types and depth of checks than gatekeeper data quality does.

Restoration data quality can be applied on batch data or on data that arrives in real time, but is limited to considering only the current set of incoming data.

Where data is RETAINED (Continuous)

Continuous data quality performs data quality using all of the data that is known and retained by the company. This makes complex entity resolution and conflict resolution possible, meaning that continuous data quality is the most powerful of the three types.

Continuous data quality can work on both batch and real-time data, and can operate in conjunction with gatekeeper and restoration data quality measures.