



cer-ti-fy sŭr'tə-fī'

1. to attest authoritatively: as a: CONFIRM b: to present in formal communication c: to attest as being true or as represented or as meeting a standard
2. to inform with certainty: ASSURE

Merriam-Webster's Collegiate Dictionary, 11th Edition

Certify Software: The Core of Data Certification

Certica Solutions gives organizations the ability to certify data in business-critical applications. The core of the company's data certification solutions is Certify™—powerful software that provides a structured, systematic approach to data quality management. With its rules-based validation engine and integrated metadata repository, Certify formalizes and centralizes the entire data certification process. The software can be applied to all types of data—in any organization or industry—across multiple database platforms.

Overview

Certify software validates data to ensure its accuracy, consistency, completeness, structural integrity and business rule compliance. The software has been designed with real-world, hands-on knowledge gained during scores of successful data quality management initiatives. Certify can be applied during data migration projects, regulatory audits, as an integral part of an inline data collection process, or as the hub for a pro-active enterprise data quality management program.

Certify allows organizations to describe the standard of optimum quality for their data, and then measure their actual data against that standard on a recurring basis. Using Certify as the locus for an application-specific or enterprise-wide data quality program gives organizations the ability to discover problems, prioritize improvements, monitor accountability and demonstrate progress over time.

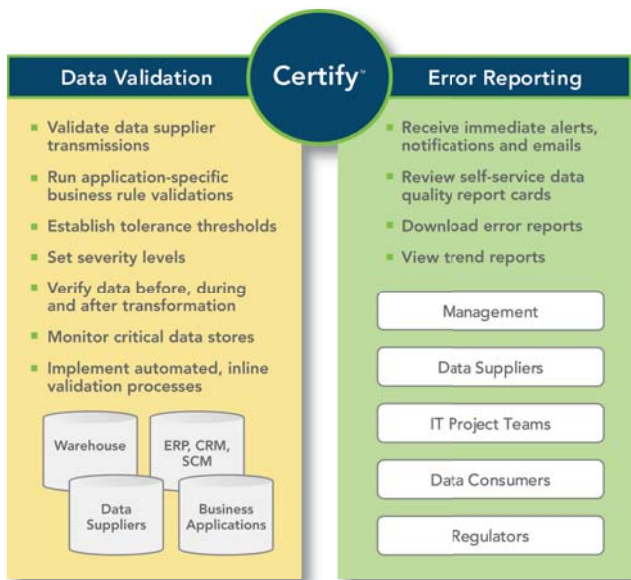
Certica Solutions emphasizes four main tenets of data quality certification, to promote data's validity and consistency:

1. **Validate incoming data at the point closest to its origin, before it is transmitted, or at least before it has had a chance to migrate to critical data stores;**
2. **Validate data quality at the table and column level, before it is consolidated, aggregated or transformed;**
3. **Use a systematic and objective approach to verify adherence to data standards and business rules; and**
4. **Make data quality certification an ongoing, repeatable process wherein trends in improvement can be observed and reported over time.**

Architecture

Certify is a web-based application that is used as an interactive console to configure and administer data certification projects, and as an inline, real-time validation web service that can be integrated into an organization's data flow.

- In an interactive mode, Certify users set up "projects" for specific data sets or databases, which include all the metadata, business rules and validations—or tests—used to examine data; and the types of results and reports they wish to see. Certify then runs tests against data on an ad hoc or periodic basis.
- Used as an inline service, Certify is triggered by another application, event or process—such as a data file being received—and automatically runs a series of pre-configured validations. Once the validations are complete, Certify can activate one or more external processes, such as a stored procedure, email notification or a data load.



Active Metadata

Certify's validation capabilities are based on metadata, business rules and Certica's data quality methodology. Metadata is actively used as the basis for data validations, and Certify users can easily update, refine and report on the metadata used in Certify projects. Certify maintains a single, consolidated repository for all metadata, as well as enterprise or application-specific business rules. When a user defines a Certify project for specified data sets or databases, metadata is automatically read from the source database schemas. The software understands which columns are primary keys, which are foreign keys and which columns are mandatory. This information is used to evaluate the structural integrity of the data.

Validation Methodology

With knowledge gained from scores of hands-on data quality improvement projects, Certica Solutions has developed a unique, six-step validation methodology which ensures complete coverage of potential data problems and promotes a formalized, repeatable process for setting up validations. Certica's data validation methodology is fully integrated into Certify, eliminating the manual coding, ad hoc queries and guesswork that would otherwise be required to set up a thorough data validation.

Certica's six-step data validation methodology is used to pinpoint data errors at the table, column and row level, providing optimum insight into the source, frequency and type of errors. The methodology can be applied to any database, at any point in the data lifecycle, in any organization or industry. The steps include:

1. Content analysis
2. Data population (completeness and validity)
3. Primary key uniqueness
4. Row uniqueness
5. Referential integrity
6. Data dependency (business rule compliance)

User Interface

Certify offers a highly intuitive, step-by-step process for configuring data validations, managing metadata and business rules, running validations against data, and viewing results.

Following are the key user interface elements which facilitate this process:

Dashboard

Once a Certify project is initiated and the software has been connected to a database, the Dashboard is used to specify and maintain metadata, such as:

- Data structures, including tables, views and user-specified queries
- Descriptions of tables and columns
- Flags indicating mandatory usage
- Primary keys
- Foreign keys
- Row uniqueness indicators

Wizard

Through its Wizard interface, Certify promotes a structured approach to data quality validation based on Certica's unique six-step data validation methodology. By moving sequentially through the steps of the methodology, Certify users ensure com-

plete coverage across the range of potential data problems—from the most simple to highly complex data errors.

Observations

Once validations have been developed, users can set up Observations, which are snapshot validations of the data at a particular point in time. Observations can be run manually or scheduled to run on a periodic basis. Observations can also be triggered whenever a specified condition is met or invoked from any web application.

Reports and Results

Certify's reporting capabilities provide rapid turn-around of results and instant notification to data users and suppliers. Results from data validations can be viewed at both the detail and summary levels, from any web browser. Data quality validations and results are easily tailored to the needs of individual users, data suppliers or business units which may need to see a particular "slice" of data. In addition, result tables can be saved for download and review.

Certify's Data Quality Report Cards allow users to interactively drill-down to data errors at the table, row and column level, to see the full instance and context of data problems. Certify offers two types of Data Quality Report Cards:

Data Quality Certification Index™ Report Card

The Data Quality Certification Index Report Card displays complete detail on the errors detected during an Observation run, identifying the exact columns and rows affected, as well as summarization based on the steps of Certica's data validation methodology, i.e., data population violations, business rule violations, etc. Certify's Data Quality Certification Index (DQCI), can be likened to a "super-metric," summarizing the overall quality of the database or data set. Certify automatically calculates the DQCI, based on the results of pre-defined data validations.



Data Quality Report Card

The Severity Level Report Card

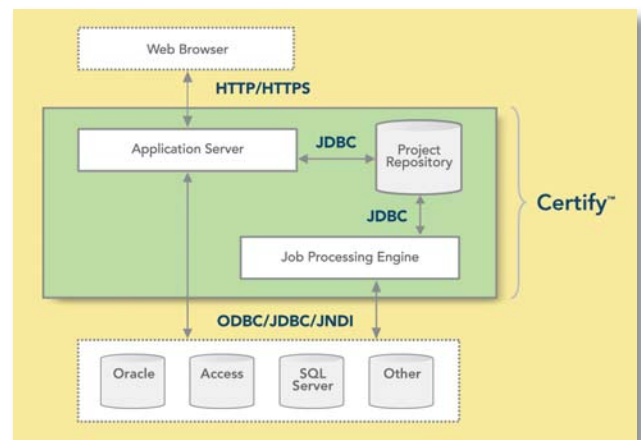
The Severity Level Report Card provides the same level of detail as the DQCI Report Card, but shows summarization based on the severity level of errors. Based on the assumption that not all data errors are equally problematic, Certify gives users the ability to easily assign a severity level to each data quality validation. Severity levels aid in the prioritization of data clean-up efforts and can flag "fatal flaw" conditions before data is allowed to populate critical databases or be reported to regulators.

Both indexes and severity levels can be used as data quality thresholds; as such, they can activate follow-on events, such as the generation of an email notification to a data supplier, the trigger of a stored procedure, or the execution of custom code to activate a subsequent process in the data flow.

Other reporting capabilities include a Trend Report which allows users to gauge the changing quality condition of data over time, and a comprehensive Metadata Report which displays all the business rules and metadata used in a Certify project.

Components

Following are the major architectural components comprised in Certify and the environmental components with which the software interacts:



Certify Components

Web Browser

Certify operates with any commercial web browser, such as Internet Explorer and Firefox, as its front-end application.

Application Server

The application server manages the user interaction with Certify by recording metadata, defining data quality validations and displaying results. Certify uses Apache Tomcat, a Java-based server, and also runs under BEA Weblogic Server from BEA Systems.

User roles, access permissions and passwords are granted and maintained by project administrators.

Database Connectivity

Certify interfaces with all ODBC-, JDBC- and JNDI-compliant databases to read source data, write result tables, and create reference tables. Certify is capable of utilizing multiple input and output databases in the same or different environments; e.g., read data from Oracle and SQL Server databases and write result tables to a MySQL database.

Job Processing Engine

The Certify job processing engine, written in Java, executes the user-defined data validations.

Project Repository

All information in a Certify project, including meta-data, test definitions and result statistics, is maintained in an integrated project repository. Certify utilizes a Firebird database as its repository and can also use either SQL Server or Oracle.

Certify Expressions

Certify incorporates its own rich inventory of expressions that support a wide range of data functions. The expressions are based on SQL, but are database independent.

Web Services

Certify's powerful web services capability allows organizations to integrate Certify as a fully-automated, inline data validation process. As such, data that is collected, extracted, migrated or reported can be automatically validated before it passes to the next step in a data flow or business process. Organizations can offer data users and data suppliers secure, remote, individualized access to Certify results, through any Internet browser.

Further, Certify can be integrated with other applications or job streams (such as a data extract-transform-load process), creating an automated, inline data validation and certification service that operates as a seamless process. Certify can be triggered by events, schedules or other applications, and can, in turn, execute stored procedures, send email notifications or trigger other applications (such as a data load).

Section 508 Compliance

Certify complies with accessibility standards under Section 508 of the Rehabilitation Act of 1973, which require that federal agencies make electronic and information technology accessible to people with disabilities.

Data Quality Analyst and Data Quality Reviewers

Certify's ability to quickly validate and report on data quality issues allows for rapid feedback between data suppliers, data consumers and data administrators. This feedback—with definitive, measurable information about the quality condition of data—promotes a valuable and lasting data quality management process, where stakeholders can be actively involved in and accountable for the improvement of data.

To achieve this rapid cycle of validation, reporting and feedback, Certify provides two main user roles: the data quality analyst and the data quality reviewer. The analyst is the "power user," responsible for configuring Certify projects, scheduling observations, creating and maintaining user profiles, etc. Reviewers are individuals—either internal or external to the organization—who review Data Quality Report Cards and other pre-defined reports online. Many times, it is the reviewers who have the responsibility for cleaning up data, especially if they are also the suppliers—or "owners"—of the data.

About Certica Solutions

Certica Solutions is a leading provider of data certification solutions. Certica helps corporate and government organizations address specific business challenges with a structured, systematic approach to identifying, reporting and monitoring data quality issues. Certica's business-specific solutions, based on the company's Certify software, enable organizations to identify costly data errors and omissions, comply confidently with regulatory requirements, improve data in a shortened time period, and certify their data suppliers' data.

©2007 Certica Solutions, Inc. Certica Solutions, Certify and the Certica Solutions logo are trademarks of Certica Solutions, Inc. All other trademarks and copyrights are property of their respective owners.



Certica Solutions, Inc.
301 Edgewater Place, Suite 110
Wakefield, Massachusetts 01880
Phone: 781.245.4515
info@certicasolutions.com
www.certicasolutions.com