

# SQL Server Best Practices for Data Quality

## Using Melissa's Data Quality Components for SSIS

### Problem:

Accurate data is imperative for an organization to conduct cost effective decision making, marketing promotions, mailings, database bloat impacting performance, storage and more. Like everything else, change is constant for your data. There is a need to cleanse and validate data when received and on a regular basis to not waste resources. Unfortunately, cleansing and validating data is difficult with the native SQL Server toolset. T-SQL, Integration Services, PowerShell and .NET all include a framework to validate strings, but significant programming logic is necessary. These technologies will not completely validate, cleanse, match and de-duplicate data. How do we leverage the SQL Server tool set to achieve these goals?

### Solution:

Let's walk through the SQL Server data quality process using these five steps to uncover the correct solution:

- SQL Server Data Import Options
- Poor Data Quality in Business is Expensive
- Correcting SQL Server Duplicates
- Resolving Out-of-Date SQL Server Data
- SQL Server Data Verification

### SQL Server Data Import Options:

SQL Server ships with a few technologies that can assist with improving the quality of your data. It is programming intensive and these technologies lack the verification and validation logic needed to deliver true data quality. These include:

- SQL Server Integration Services
- BIML
- T-SQL
- PowerShell
- .NET
- Power BI

SQL Server Professionals are familiar with SQL Server Integration Services (SSIS) as a means to perform extraction, transformation and loading (ETL) operations. SSIS is generally the main tool used by SQL Server Professionals to execute ETL processes with interfaces to numerous database platforms, flat files, Excel, etc. SSIS also includes the ability to include T-SQL and .NET scripting logic all in a graphical interface. Although all organizations are different and each project has different requirements, SSIS is generally the tool selected when importing and exporting data. To help automate the development of SSIS Packages, BIML can also be used to programmatically build SSIS Packages to include the control flow, data flow, connection managers and error handling.

The T-SQL language with the INSERT, INSERT... SELECT, BCP, BULK INSERT, OPENROWSET, etc. commands enable SQL Server Developers to perform many of the same tasks in SSIS. The same is true for PowerShell and .NET. There are libraries available that are able to extract, load and transform data, but all are programmatic. At times directly programming in T-SQL, PowerShell or .NET yields higher performing code, but comes at a high cost for development and overall maintenance.

Power BI also has a means to import and transform data within a report for presentation purposes, but this technique is generally beneficial for a single report and does not validate the accuracy of the data set.

The technology teams and data stewards need to work with a familiar tool to perform ETL operations while simultaneously cleansing the data.

### Poor Data Quality in Business is Expensive

“Garbage in yields garbage out.” “You Reap What You Sow.” Regardless of the analogy, your decision making is only as good as the quality of your data. By some accounts, 25% of data is out of date 1 year after collected, making data validation a recurring process rather than a one-time event. Unfortunately, verifying data is very difficult with numerous internal and external applications that interact with customers, suppliers and partners. Data does not necessarily enter the organization from a single point of entry and data gets out of sync between systems when customers move, change contact information, make career advancements, have family changes, etc. and all impact the accuracy of your data.

Beyond the tangible costs associated with poor data related to marketing promotions, mailings and more, inaccurate data results in time and budget waste from the significant time to process the error related data and/or promotions never reaching the intended audience. On the technology side, inaccurate data bloats databases impeding performance, extending backup times and maintenance windows which ultimately impacts the operating hours for the organization. Organizations need a means to validate and correct their data to ensure it is accurate.

### Correcting SQL Server Duplicates

When trying to validate data, there is nothing more frustrating than trying to determine if “John Smith” is the same person or if they have been entered into the system 10 different ways. What is being referred to here is not literally string matches where we can write T-SQL code to compare “John Smith” vs. “John Smith”, it is a matter of determining if multiple people are the same person based on various moves, contact information and more. Often times employees are afraid to delete data to remove potential duplicate records, but this comes as an expense to the organization when mailings are returned and time is wasted contacting the wrong customer. What is really needed is a means to validate and consolidate duplicate customer records with confidence.

### Resolving Out-of-Date SQL Server Data

Change is constant; this is the case with your data as well. Old and stale data can be considered no longer relevant, but is it? Historical customers already familiar with your brand and the quality of your product may have fallen through the cracks based on moves and life changes. If this out-of-date data is the roadblock to re-engaging with these prior customers, cleansing and enhancing the data can have a direct impact on the bottom line for the organization. Turn your data into a goldmine as a means to re-engage and retain your customers.

Organizations need a simple means to regularly validate their data which traditionally is very time consuming and resource intensive. How can organizations make this a competitive advantage?

## SQL Server Data Verification

Data is one of the most valuable assets for organizations, but your data needs to be designed, developed and maintained just as a core business product. Organizations need a simple means to validate, cleanse, match and de-duplicate data. One solution that I would like to recommend is [Melissa's Data Quality Components for SSIS](#) Melissa's suite of tools to address critical data quality issues:

- Melissa Contact Verify
- Melissa Global Verify
- Melissa Matchup
- Melissa Fuzzy Match
- Melissa Personator
- and more...

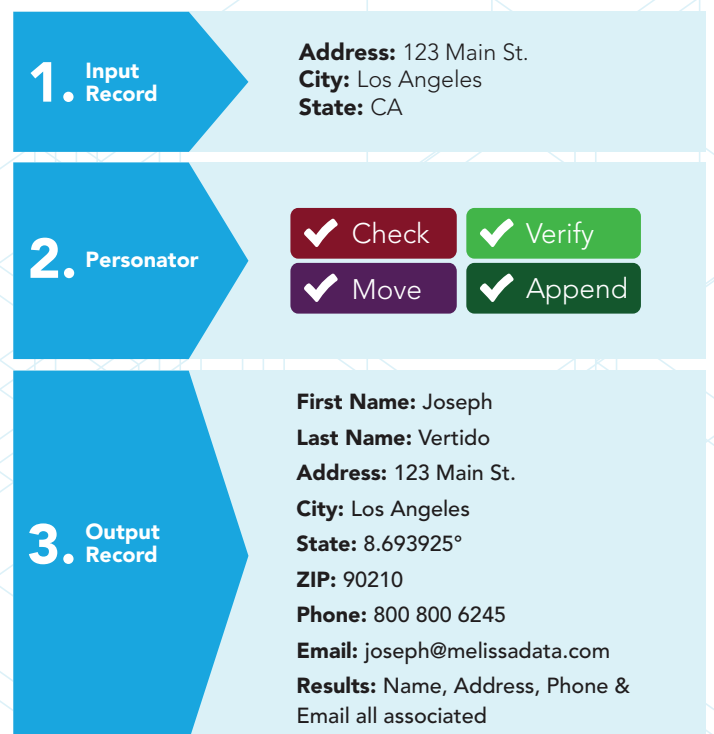
Learn how Melissa's solutions address critical business challenges:

Business Challenge	Melissa Solution
<b>Importing Data</b>	Melissa's solution directly integrates with SSIS which is the tool of choice for SQL Server Business Intelligence Professionals, Developers and DBAs for both on-premises and cloud-based ELT projects.
<b>Poor Data Quality</b>	Melissa's Contact Verify and Global Verify are fundamental components to validate name, address, email, phone and more data in the United States and globally for 200+ countries. This enables you to ensure the accuracy of your data directly in SQL Server Integration Services.
<b>Correct Duplicate Data</b>	Melissa's Matchup component is able to confidently de-duplicate data providing a final golden record with a data quality score.
<b>Enrich Customer Records</b>	Melissa's Personator, BusinessCoder, Property and IP Location components can add myriad data elements to records including latitude/longitude coordinates and location intelligence, property characteristics and mortgage details, demographic variables and business firmographic data including SIC codes, sales volume and employee size.
<b>Resolve Outdated Data</b>	Beyond the components already mentioned, Melissa's Fuzzy Match is based on a state-of-the-art matching algorithm to validate and rank your data.

Business Challenge	Melissa Solution
<b>Data Verification</b>	With each Melissa component, the final data is accompanied by result codes specifying the actions taken on the field and enables you to pass the updated data between components to have the best final data set.
<b>Recurring Data Validation</b>	Melissa offers two options to validate your data on a recurring basis. You can use the default behavior in SSIS with calls out to Melissa's current data set or quarterly updates are available to use when all data must remain behind the firewall.
<b>Low Learning Curve</b>	Drag and Drop Melissa components from the SSIS Toolbox directly into the Data Flow Tasks of existing packages to improve the data quality of your projects. This yields a very low learning curve for SQL Server Professionals enabling them to quickly integrate the Melissa components and demonstrate value to the organization.

## Sample Melissa Solution

As an example, when working with Melissa's Personator component for SSIS, your existing data is used as an input. As shown in the image below, you can see a minimal amount of data is available as the input. With integration with the Personator Melissa component the data is checked, verified and appended. The output from the process is a complete record with name, address and contact information. This enables the organization to better utilize the data for marketing, mailing promotions and more.

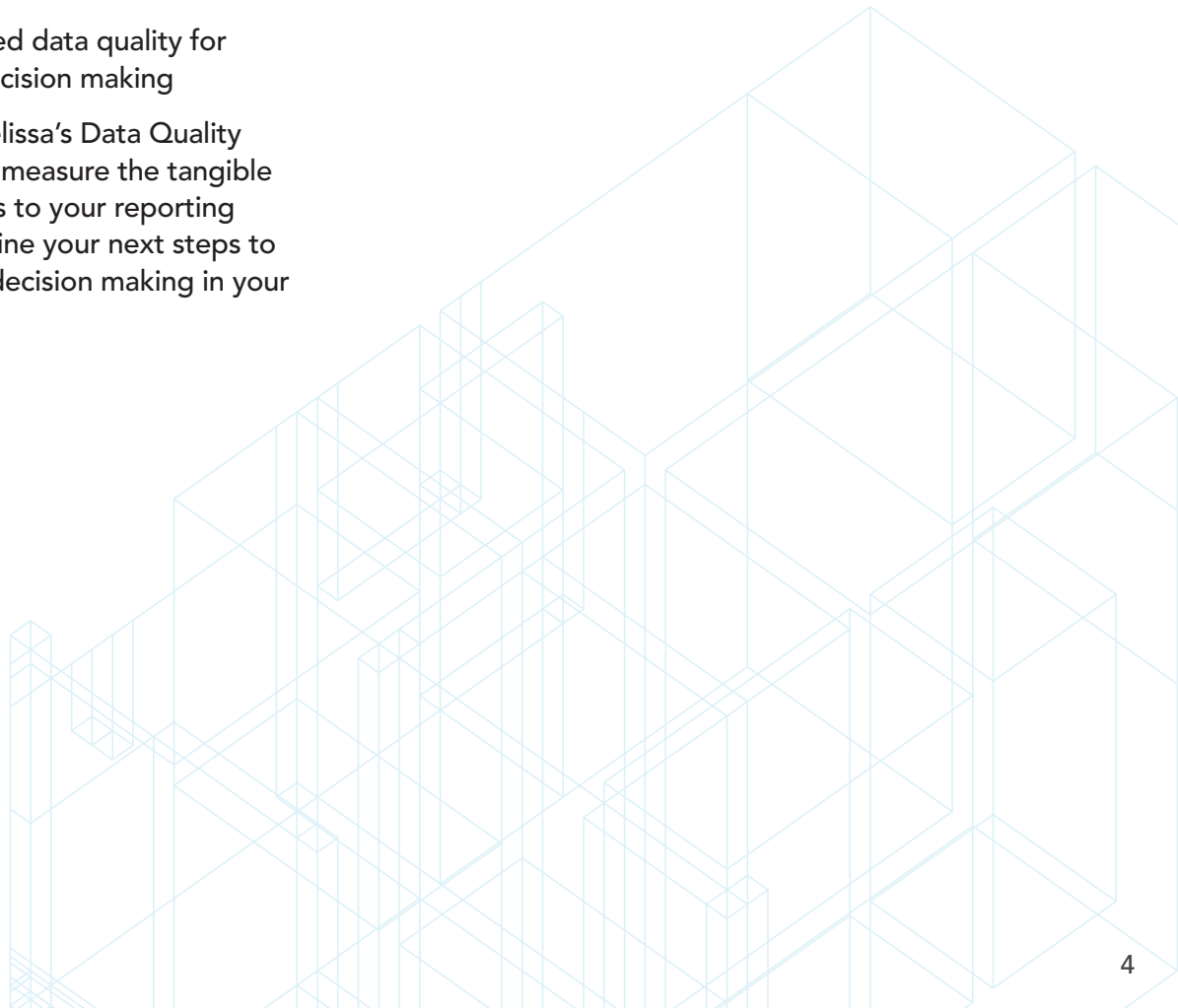


## How do I get started with the Melissa Data Quality Components for SSIS?

- Learn more about all Melissa's Data Quality Components for SSIS and what they will do to help your organization.
- Download your personal evaluation version of Melissa's Data Quality Components for SSIS.
- Schedule a personal demo of Melissa's Data Quality Components for SSIS with a member of the Melissa team and then put the tool through its paces in your environment with your data. See how easy it is to improve the data quality at your organization.
- Communicate how Melissa's Data Quality Components for SSIS can help your organization:
  - Time savings for data validation and cleansing
  - Seamless integration into existing SSIS Processes
  - Significant functionality out of the box with the ability to customize as needed
  - Radically improved data quality for reporting and decision making
- After working with Melissa's Data Quality Components for SSIS, measure the tangible and intangible benefits to your reporting processes and determine your next steps to radically improve the decision making in your organization.

## Next Steps

- Learn about the value of Melissa's Data Quality Components for SSIS.
- Get a personal demo of Melissa's Data Quality Components for SSIS.
- Download Melissa's Data Quality Components for SSIS.
- Check out these videos:
  - Data Quality for SSIS Video
  - Fundamentals of SQL Server Contact Data Cleansing





### [About the author](#)

Since 2002, Jeremy Kadlec has delivered value to the global SQL Server community as an MSSQLTips.com co-founder and Edgewood Solutions SQL Server Consultant.

**Original Source** - <https://www.mssqltips.com/sqlservertip/6210/sql-server-best-practices-for-data-quality/>

## About MSSQLTips.com

MSSQLTips.com is a free community dedicated to SQL Server. Since 2006 our team has delivered value to millions of SQL Server DBAs, Developers and Business Intelligence Professionals on a daily basis. MSSQLTips.com helps solve real world problems and improve your SQL Server knowledge with free tips, tutorials, web casts, videos and more. The MSSQLTips.com team is comprised of **150+ expert authors** led by **Greg Robidoux** and **Jeremy Kadlec**. We are focused on sharing knowledge and improving the global SQL Server community every second of the day.

Licensed with Permission to Melissa from Edgewood Solutions, LLC.

Copyright (c) 2006-2020 [Edgewood Solutions, LLC](#) All rights reserved

## ABOUT MELISSA

Melissa is a leading provider of data quality, identity verification and address management solutions. Melissa helps businesses win and retain customers, validate and correct contact details, optimize their marketing ROI and manage risk. Since 1985, Melissa has been a trusted partner for key industries like retail, education, healthcare, insurance, finance, and government. For more information, visit [www.melissa.com](http://www.melissa.com) or call 1-800-Melissa.