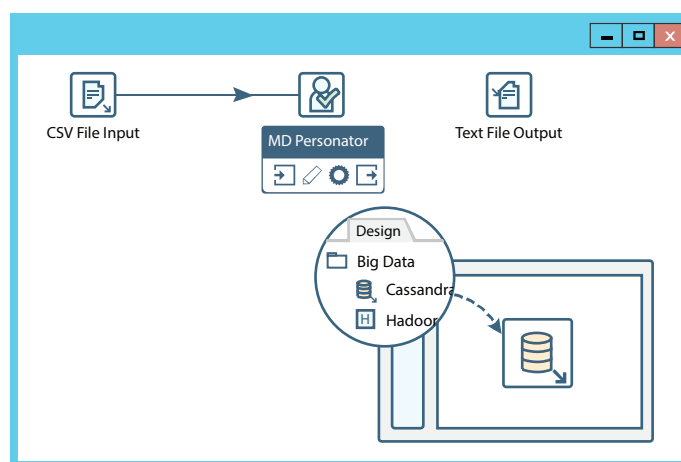# Data Quality Components
for Pentaho®

Turn Big Data into Trusted, Actionable Intelligence

Melissa's Data Quality Components for Pentaho give companies a competitive advantage by leveraging Customer Data Integration (CDI) to transform Big Data into useful, knowledgeable information. Get better analytics, improve sales and marketing campaigns, and optimize your data warehousing efforts with clean, enriched, and up-to-date contact data. Migrate and clean data at the same time. That's the power of adding Melissa Pentaho components.

Robust ETL & Customer Data Integration

Data Quality Components for Pentaho® reads from and writes to relational databases (SQL Server®, Oracle®, Netezza®, DB2®, etc.), fixed or delimited text files, XML, JSON, COBOL, and other file formats such as Avro® and Parquet®, or Hadoop®-based NoSQL stores such as HBase® and Hive®. Go beyond standard ETL tools to scalable and flexible management of end-to-end data flows.

Data Quality Components for Pentaho's graphical user interface and re-usable macros and operations make doing work quick and easy.

Full Spectrum Data Quality

Melissa's integration with Pentaho leverages the full spectrum of Melissa's data solutions, so you can:

PROFILE DATA to identify weak points in the data collectionprocess.

CLEANSE DATA with a variety of programmatic or regular expressions to correct data inconsistencies.

VERIFY DATA using powerful global address, name, phone, and email validation processes, and match name-to-address to verify ID.

ENRICH DATA by adding valuable demographics, IP location, and property data for greater insights.

UPDATE DATA by accessing up to 20 years of change-of-address records to get the most current address of your customers.

MATCH & MERGE DATA with advanced fuzzy matching algorithms and survivorship rules to eliminate duplicates.

MONITOR DATA over time to enforce business rules on incoming records and reduce the amount of cleansing needed.

✓ 📁 Input

CSV file input	Generate random credit card n	Get repository names	Load file content in memory
Data Grid	Generate random value	Get table names	Microsoft Access Input
De-serialize from file	Get File Names	Google Analytics	Microsoft Excel Input
ESRI Shapefile Reader	Get Files Rows Count	HL7 Input	Mondrian Input
Email messages input	Get SubFolder names	JSON Input	OLAP Input
Fixed file input	Get System Info	LDAP Input	Property Input
GZIP CSV Input	Get data from XML	LDIF Input	RSS Input
Generate Rows			

Data Quality Components for Pentaho reads from and writes to many relational databases and multiple file formats.

melissa

Melissa Components Available for Pentaho:

Profiling
- Identify data that fails to comply with business rules & standards (discover database problems and fix them)

Personator®
- Verify name corresponds to address, email, and phone data to identify customers' true identities
- Append missing name, phone, and email address
- Enhance data with valuable demographic info like household income, occupation, marital status, and more

Global Contact Verify
- Verify, correct, and standardize addresses, phones, and emails for U.S. and Canada, plus over 240 countries (including real-time email mailbox validation)
- Geocode addresses by adding rooftop lat/long coordinates
- Transliterate 8 major alphabets and display in native or Roman characters
- Enrich U.S. data with census information (congressional district, county, time zone, and more)

General Cleansing
- Clean, standardize, and reformat any data type: change casing or capitalization; add/remove punctuation; expand/contract abbreviations
- Search and replace parts of a string
- Create your own rules (triggering) to standardize any data type

SmartMover℠
- U.S./Canadian change-of-address processing to update addresses of people or businesses that have moved

MatchUp®
- Dedupe records with domain-specific knowledge of contact information
- Match records with data inconsistencies

IP Locator
- Identify web visitor's geographic location

BusinessCoder®
- Enrich contacts with valuable firmographics from 25 million multisourced business records
- Over 25 return fields such as StockTicker, phone numbers, employee size, and sales estimates

Property
- Enrich contacts with comprehensive property and mortgage data from over 140 million records
- Add information from over 160 fields such as parcel details, prior sales, owners, and more

---

Melissa
22382 Avenida Empresa RSM, CA 92688-2112 | 1-800-MELISSA | www.melissa.com

Find us on: f G+ in t YouTube