

Datagaps ETL Validator

Quick Start Guide

A highly effective tool for automating and easing the testing process for data integration, data migration, and data warehouse / BI projects.

ETL Validator is a comprehensive solution for automating the testing of data integration, data migration, and data warehouse projects. It makes use of Datagaps patented ELV architecture (Extract, Load and Validate) to extract test data from heterogeneous data sources, load it to a test data store and run validations on the test data set.

Datagaps offers a comprehensive regression and performance testing tool called BI Validator for testing BI applications built using OBIEE, Tableau, SAP Business Objects and IBM Cognos.

This Help documentation is designed so you can quickly learn ETL Validator as a new user or enhance your knowledge as a regular user.



What ETL Validator Can Do for You

- Compare and report results from data integration and data warehouse source to target ETL data loads
- Create test case repositories with multiple test cases to be used for periodic and regression tests
- Provides for definition of data source and target model rules to support data profiling and data validity testing
- Provides facilities for testing relational database (RDBMS) referential integrity (i.e., primary and foreign key integrity among tables), in addition to flat file and big data test capabilities
- Compare data file and database metadata (e.g., development vs. test) when they should be the same
- Help create test cases with a SQL query builder for multiple source to single target test verifications

Get started in 6 easy steps

1 - CREATE CONNECTIONS TO YOUR DATA FILES AND DATABASES

Creating connections is the first step to using ETL Validator. A new connection can be created by clicking on the **Connections** button from the 'Home' page. ETL Validator supports connections to most popular relational databases as well as flat files, XML files and Excel files.

Files local to Server	Relational
Flat File (CSV or Delimited)	Oracle
XML File	Microsoft SQL Server
Excel File	IBM DB2 for LUW
	IBM DB2 for iSeries
Flat files at SFTP location	IBM DB2 for z/OS
SFTP	IBM PDA (Netezza)
	PostgreSQL
Big Data	Pivotal Greenplum
Hadoop Hive 2 (and Impala)	MySQL
Apache Drill (parquet, avro, json)	HP Vertica
Spark SQL (coming soon)	SAP Sybase
	Teradata
Cloud	NoSQL
Salesforce	Datastax Enterprise (Cassandra)
Amazon S3 (using Drill)	MongoDB (using Drill)
Amazon Redshift (coming soon)	Hbase (using Drill)

ETL Validator New Connections Window

The example above is for an Oracle DB. The window options may be different for other DB connections (e.g., DB2, MySQL).

Name: data connection name of your choice

Data Source: type of data source (select from dropdown list)

User Name: name of the database user in the database

Password: password for the database user

Host Name: the server hosting the database

Port: port number of the database listener. For example, default port for Oracle database is 1521

Test the new connection by clicking “Test”, then Save. The ETL Validator Connections window will now be displayed with your new connections.

ETL Validator Connections Window

Name	Type	User	Connection Typ	Created Date	Data Model	Status
Data Source						
FF_Connection	Flat File / Excel		Data Source	12/2/2015 2:28:07 PM	Create	Connected
Sample	PostgreSQL	postgres	Data Source	2/13/2015 6:08:16 PM	View / Update	Connected
Wayne Flat-source	Flat File / Excel		Data Source	12/6/2016 1:24:13 AM	Create	Connected
Wayne Flat-target	Flat File / Excel		Data Source	12/5/2016 8:44:57 PM	Create	Connected
Workschema						
Workschema	PostgreSQL	postgres	Workschema		View / Update	Connected

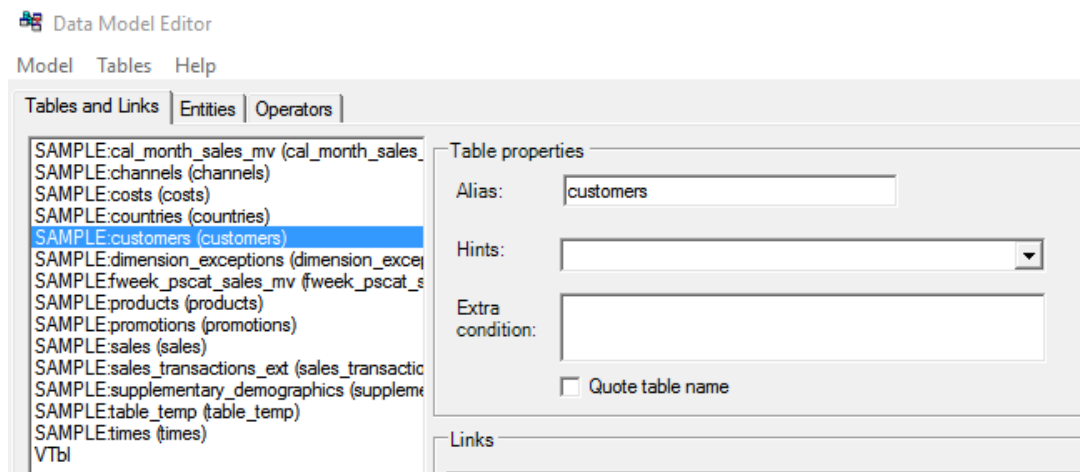
2 - CREATE A DATA MODEL FOR EACH OF YOUR DB CONNECTIONS

Every database (DB) connection (i.e., Oracle, DB2, but not flat file or XML file) you create must be represented by an ETL Validator “Data Model. Information to complete the data model can usually be found in your ER data models, project data mapping documents, test plans, or data management tool /editor (e.g., SSMS, Toad). The data model is used to link related tables, and provides individual table, entity, and attribute specifications which

will be used to build ETL Validator queries. In addition, users can specify operators like comparisons to be used in queries. For more information, see the Data Model Editor User Guide delivered with the product.

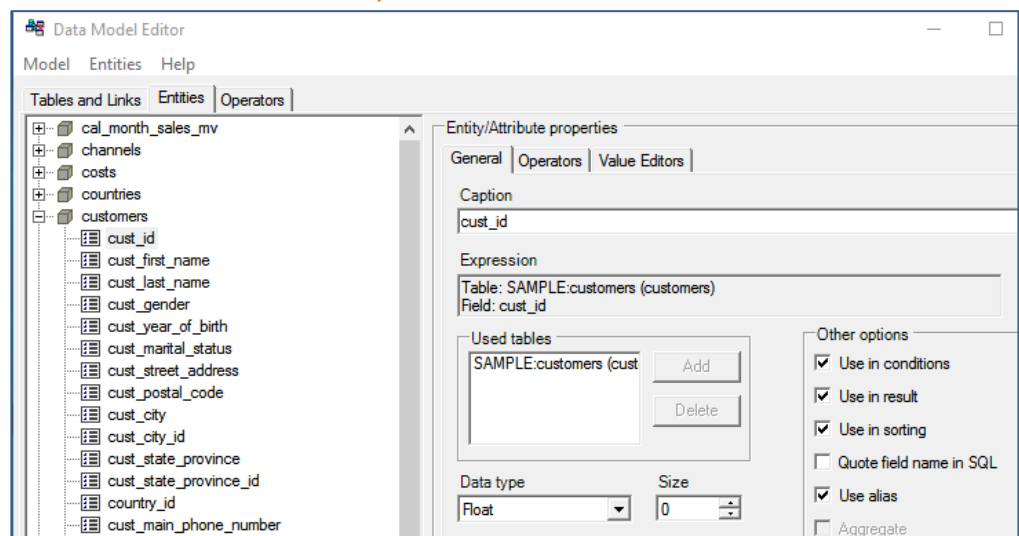
1. Click **Connections** in the ETL Validator Home Page.
2. In the **Data Model** field, select **Create** (or **View/Update** if you are updating a data model) in the row associated with the database for which a data model is needed.
3. A **Schema Selection** window will be displayed. From that window, select the schema. For each table to be used for testing source or target values, select **Add**. Then select **Entity Attribute Properties** as appropriate for tests you plan.
4. Click **Entities** tab. For each column you plan to include in ETL Validator tests, select appropriate attributes properties under the **General**, **Operators**, and **Value Editors** tabs.

ETL Validator Data Model Definition Window



Select tables one at a time to specify alias names and links to other tables to be used during testing.

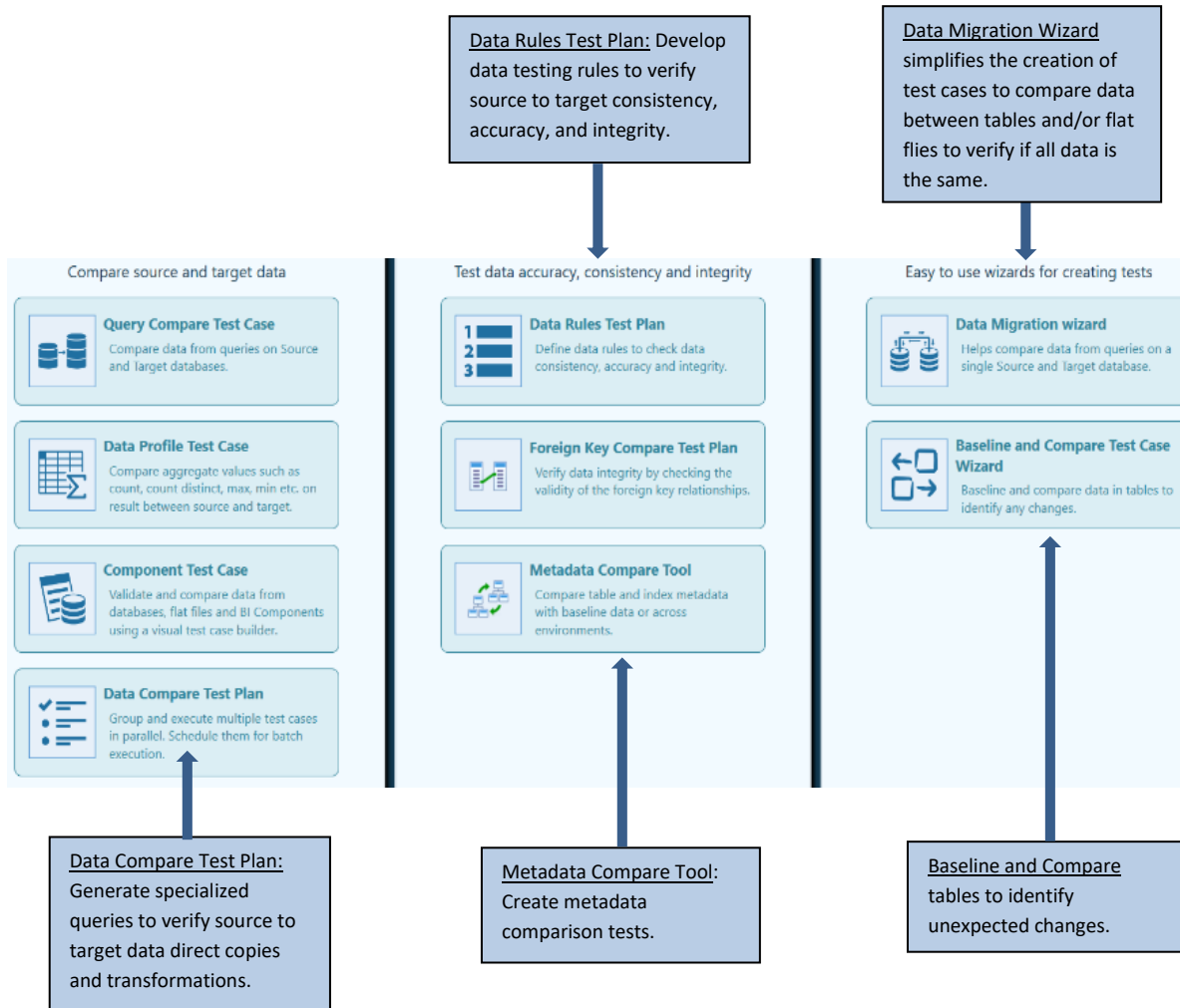
ETL Validator Data Model Entity and Attribute Definition Window



For each table to be used in testing, select table attributes which will be used during ETL Validator runs. See the ETL Validator Data Model User Guide for complete information.

3 – PLAN YOUR ETL VALIDATOR TESTING

The Main Menu contains most of the ETL Validator functions you will need to fulfil the objectives of your ETL test plans. Here, we point out some of those that you are most likely to use.



Listed below are a few of many test conditions that are supported with the variety of ETL Validator functions listed above from the ETL Validator Main Menu.

Query Compare Test Cases

- Create query test cases that compare and verify data copied directly (no transformations) from source to target and that no records or columns were dropped.
- Develop tests to verify that data values were not truncated during ETL from source to target.

Data Profile Test Cases

- Develop tests that compare aggregate values, distinct values, min/max values between source and targets.
- Verify that data was completely loaded to targets
- Verify that “not null” fields were populated as expected.

Component Test Cases

- Create tests that check for duplicate records and duplicate values in target data
- Verify that surrogate keys uniquely identify rows of data

- Test that cleansing and transformations of data from source to target meet business requirements.
- Write tests that use ETL Validator “Lookup” components to verify that source to target lookups were processed correctly.
- Develop tests to assure that numeric value precisions are as required in target data

Data Compare Test Plan

- Develop tests that verify all tables and specified columns were loaded from source to target correctly.

Foreign Key Compare Test Plan

- Develop tests to verify referential integrity of joins between tables (ex. primary and foreign keys). This test represents the ability to detect orphan records based on primary-foreign key relationships.

Metadata Compare Test Tool

- Generate tests that verify table and index data values are propagated across all environments (eg., development, QA, UAT)

Data Migration Wizard

- Generate tests to compare between sources and targets when data has been directly copied during ETL’s (no cleansing, no transformations).

Baseline Compare Test Case Wizard

- Develop tests that may be run repetitively to verify no regression or unexpected changes in the target.

4 – CREATE TEST CASES USING ETL VALIDATOR COMPONENTS

Each “Component Test Case” you create utilizes the Datagaps ELV architecture to extract data from your data model definitions such as databases, flat files, XML, etc. in order to compare and assess whether they meet your test specifications.

Component Test Case provides a visual test case builder for defining database tables / flat files, etc. and validations. Some of the common scenarios where the Component Test Case can be used are listed below

- Compare data in a flat file with the database table
- Compare data from multiple data types and source to a target
- Test incremental data loads to assure new data meets requirements
- Regression test a new ETL load by comparing baselined data with the latest load
- Test complex transformations

An example is explained in a short online ETL Validator video:

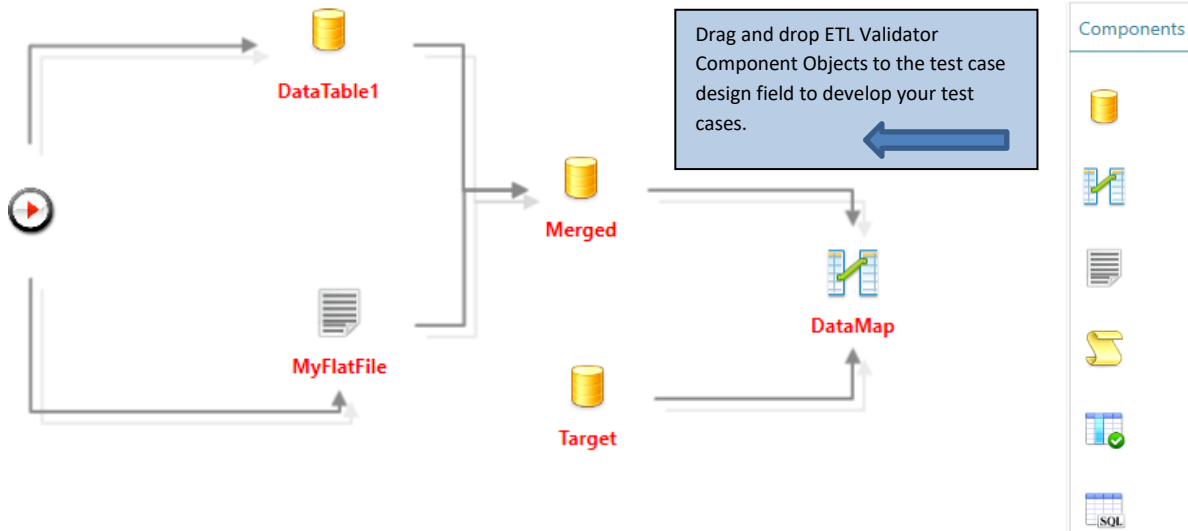
<http://datagaps.freshdesk.com/support/solutions/articles/1000238879-advanced-etl-testing-using-a-component-test-case>

The The ETL Validator “Component Test Case” builder uses drag and drop features to aid in building test cases.

To understand the capabilities described in the video, see the diagram below. In this example, a table (DataTable1) and a flat file (MyFlatFile) contain related information. When the test case is run, data from each data source will be extracted and merged into one file (Merged) to represent the actual ETL to be tested. Using the ETL Validator mapping function, the user creates column join conditions (DataMap) for the Merged and Target components. At this point, the test case can be verified, saved, then run.

ETL Validator creates a detailed report after the test is run to provide you with test results.

ETL Validator Component Building Using Drag and Drop



5 – CREATE TEST PLANS TO GROUP TEST CASES

Test Plans are an important ETL Validator feature which can be used to group ETL Validator test cases and run them. ETL Validator provides for different types of Test Plans and those can be used based on user requirement. Many of these are supported by “wizards” for ease of development.

The types of Test Plans:

- Data Compare
- Data Rules
- Foreign Key

ETL Validator Test Plans Repository

Test Plans				
<div>Test Plans</div> <ul style="list-style-type: none"> Data Compare Data Rules Foreign Key Compare 	Search : <input type="text" value="Search text here"/>		Search by label : <input type="text"/>	
	<input type="checkbox"/>	Name	Test Plan Type	Test Suite
	Data Rules			
	<input type="checkbox"/>	Data Rules_5459	Data Rules	Test Plans
	<input type="checkbox"/>	Data Rules_3382	Data Rules	Test Plans
	<input type="checkbox"/>	Data Rules_1411	Data Rules	Test Plans
	<input type="checkbox"/>	Data Rules_8275	Data Rules	Test Plans
	Data Compare			
	<input type="checkbox"/>	Data Compare_8727	Data Compare	Test Plans
	<input type="checkbox"/>	Data Compare_1229	Data Compare	Test Plans

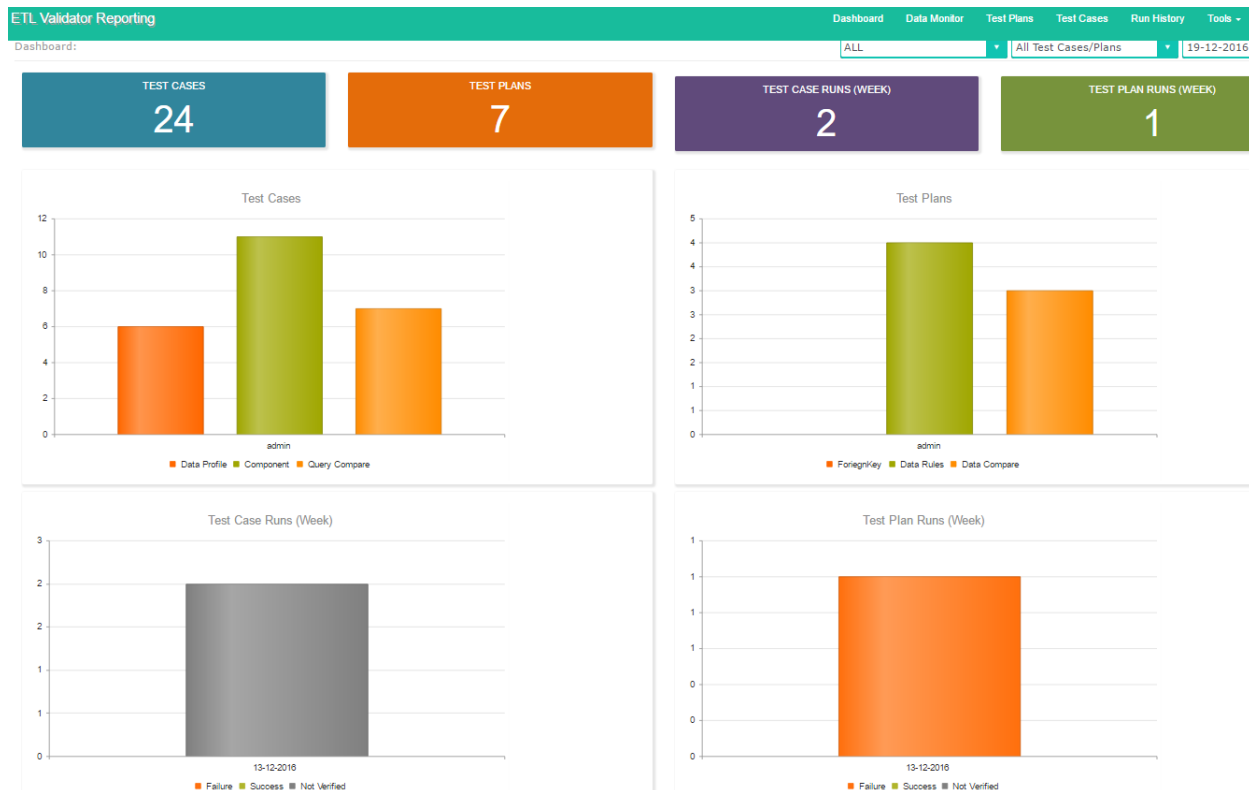
6 – PLAN ETL VALIDATOR REPORTS TO DISPLAY TEST RESULTS

As you create and run tests, you want the ability to track your status. For example, how many test plans and test cases do you have and how many in each category were run.

What reporting options do you have with ETL Validator?

- **Dashboard** with run metrics: totals of test cases and plans run, failed, passed.
- **Data Monitor** which allows for selective reporting by user chosen test plans and test cases.
- **Test Run History** which provides a history of runs by selected test plans and test cases.

ETL Validator Test Reporting Dashboard



ETL Validator Testing History Report

This chart lists test results by **Test Plans** and **Test Cases** with *Run Status*, *Start time*, *Run Duration*, and *Run Result Messages*.

ETL Validator Reporting

Dashboard

Run History:

Name	Run Status	Start Time	Run Duration(...)	Message
▼ runtime: Test Cases				
QC_Sample	FAILURE	2015-12-02 13:32:27.286	1	Test Case Run Completed Successfully
Sample_DB_FF	FAILURE	2015-12-02 14:44:57.293	11	Test Case Run Completed Successfully
Component Test Case_2426WY	FAILURE	2016-12-07 15:26:42.574	0	
Data Profile TC_6490	NOT VERIFIED	2016-12-13 10:01:25.663	0	Unable to execute the test case. Please verify.
Component Test Case_4040	NOT VERIFIED	2016-12-13 10:01:25.686	0	Unable to execute the test case. Please verify.
▼ runtime: Test Plans				
Data Compare_1229	FAILURE	2016-12-13 10:01:25.552	5	Completed but one or more Test cases failed

ETL VALIDATOR SUPPORT: User Guide, Use Cases, Videos, Blogs

The Datagaps team knows you want a variety of documentation to lead you through effective usage. We suggest first starting with this Quick Start Guide which will guide through the creation of 1) Connections to data, 2) Data Models defining the types of tests you need for each connected data source, then 3) creating test cases. Lastly, verify and run each test case to assure they are working correctly.

During your product start up process, use the [ETL Validator User and Reference Guide](#) (access through the product Help dropdown), [videos](#) available at www.Youtube.com/datagaps, and [ETL Validator Use Cases](#) available on the datagaps website, www.datagaps.com. You will find these materials to be very useful.

ETL VALIDATOR TERMINOLOGY AND DEFINITIONS

TERM	DEFINITION
Baseline and Compare	This wizard is used to define comparisons between tables to identify unexpected changes.
Component	Components are a basic unit created by users to build an ETL Validator test case. When users run a test case with a “component” they defined, the ETL Validator server executes SQL queries and other functions to get data from the database or flat files specified in the database connection. Components can be database, flat file, data profile, mapping, data rule, etc.)
Component Test Cases	Component Test Cases support comparison or validation of data from multiple heterogeneous sources. They are built from “components” created by ETL Validator users.
Data Compare Test Plan	Used to group test cases together allowing them to be run in parallel or sequentially.
Data Model	An ETL Validator data model is created by users for each database to define rules for testing each table and column.
Data Rules Test Plan	An ETL Validator Data Rules Test Plan is created by users to define rules that will be applied to each source and target table while testing data consistency, accuracy and integrity. The rules defined for the tables and columns are based on the data model associated with each connection.
Data Source Connection	Each database referenced (both source and target) in ETL Validator tests must be connected using the ETL Validator function.
FF Component	A flat file test definition created by users to identify the testing to be conducted
Lookup	An ETL Validator feature to add generic Lookups so user can use them to build data rules for different components.
Mapping compare	Mapping Compare Component is a basic unit used to build a test case. It can be used to compare and validate data from other user developed ETL Validator components. It accepts two input connections from any of the DB/FF components.
Metadata Compare	The metadata wizard allows users to compare table and index metadata with baseline data or across environments.
Query Compare	Query Compare Test Cases are used to compare data from source and target queries to determine if the data in source and target are the same.
Sources	A production database table or flat or XML files, etc., that feed data through ETL’s into a target database; used as input to ETL Validator.
Source Systems	A term used by Datagaps referring to all data (databases, tables, flat files) that are input to ETL Validator for testing in its test plans and test cases.
Targets	The database(s) in which data will be loaded or inserted by means of ETL from source data; used as input to ETL Validator.
WorkSchema	The ETL Validator Workschema is where this program stores temporary tables with test data results from the source systems. For instance, test data can be the output of a query from a DB Component or data loaded from a Flat file in a Flat file component. When a DB Component is baselined, a copy of the data from the database query is also saved in the benchmark table in the workschema. The Workschema also has the validation results and the sample comparison data.

NEED MORE HELLP?

TRAINING RESOURCES

Whether you want to get the basics or fine-tune your skills, Datagaps can help with a variety of targeted training opportunities that include recorded trainings, live expert-led WebEx sessions, and downloadable materials at Datagaps.com

Subscribe to our YouTube channel to receive training video updates at youtube.com/datagaps

TECHNICAL HELP

Search our Frequently Asked Questions Knowledgebase or contact our expert technical support staff at Datagaps.com/support

CONTACT SALES

Phone: 1- (571) 305-2345

Email: contact@datagaps.com