



Data Integrator Release Summary

Data Integrator 11.5.1

for Windows and UNIX

Copyright

If you find any problems with this documentation, please report them to Business Objects S.A. in writing at documentation@businessobjects.com.

Copyright © Business Objects S.A. 2004. All rights reserved.

Trademarks

Business Objects, the Business Objects logo, Crystal Reports, and Crystal Enterprise are trademarks or registered trademarks of Business Objects SA or its affiliated companies in the United States and other countries. All other names mentioned herein may be trademarks of their respective owners.

Third-party contributors

Business Objects products in this release may contain redistributions of software licensed from third-party contributors. Some of these individual components may also be available under alternative licenses. A partial listing of third-party contributors that have requested or permitted acknowledgments, as well as required notices, can be found at:

<http://www.businessobjects.com/thirdparty>

Patents

Business Objects owns the following U.S. patents, which may cover products that are offered and sold by Business Objects: 5,555,403, 6,247,008 B1, 6,578,027 B2, 6,490,593 and 6,289,352.

Date

December 16, 2005



Contents

Introduction	5
Overview	5
New to Data Integrator version 11.5.1	5
New to Data Integrator version 11.5.0	6
Trusted information	6
Case preservation for database object names	7
Data Quality dashboard metadata reports	7
Data profiler redesign	7
End-to-end metadata viewing from Desktop Intelligence documents ..	8
Validation transform enhancements	9
Variable-length character processing enhancements	10
Productivity	10
BusinessObjects™ Composer	10
COBOL copybook file format enhancements	11
Function enhancement (rand_ext)	12
Metadata reports redesign	12
Impact and Lineage	12
Operational Dashboards	12
Data Quality dashboards	13
Auto Documentation	13
Query transform enhancements	13
Tutorial upgrade	13
Web Services option enhancement	14
Windows clustering failover support	14
Scalability	14
Changed-data capture (CDC) enhancement	15

Contents

Oracle Real Application Cluster (RAC) support 15

Performance improvements 15

Teradata throughput improvement 16

XML_Pipeline transform 16

Introduction

Welcome to BusinessObjects Data Integrator XI Release 2 version 11.5.1. This document summarizes the new features for Data Integrator versions 11.5.0 and 11.5.1. It also describes a new product that is complementary to Data Integrator—BusinessObjects™ Composer.

For important information about this product release including installation notes, resolved issues, and known issues, see the *Data Integrator Release Notes*.

For details on all features, see the *Data Integrator Technical Manuals*, which is available from the Data Integrator Designer Help menu, the Doc\Books directory in your Data Integrator installation, or from the Windows Start > Programs menu.

For additional information that might have become available after the publication of this summary, visit the Business Objects documentation Web site at http://support.businessobjects.com/documentation/product_guides/default.asp.

Overview

The features presented in Data Integrator XI Release 2 accommodate several key areas:

- [Trusted information](#)
- [Productivity](#)
- [Scalability](#)

Each new feature (or product) supports a key area. Find new feature and product summary information under the associated key area heading.

The following lists itemize new features by version, then alphabetically.

New to Data Integrator version 11.5.1

The following features are new to Data Integrator version 11.5.1:

- [Case preservation for database object names](#)
- [Changed-data capture \(CDC\) enhancement](#)
- [Data Quality dashboards](#) (new module in metadata reports application)
- [Oracle Real Application Cluster \(RAC\) support](#)
- [Query transform enhancements](#)

- [Tutorial upgrade](#)
- [Validation transform enhancements](#)

New to Data Integrator version 11.5.0

The following features are new to Data Integrator version 11.5.0:

- [BusinessObjects™ Composer](#)
- [COBOL copybook file format enhancements](#)
- [Data profiler redesign](#)
- [End-to-end metadata viewing from Desktop Intelligence documents](#)
- [Function enhancement \(rand_ext\)](#)
- [Metadata reports redesign](#) (including Impact and Lineage, Operational Dashboards, and Auto Documentation modules)
- [Performance improvements](#)
- [Teradata throughput improvement](#)
- [Variable-length character processing enhancements](#)
- [Web Services option enhancement](#)
- [Windows clustering failover support](#)
- [XML_Pipeline transform](#)

Trusted information

This version of Data Integrator provides additional features that help you verify the origin, quality, and integrity of the data in your projects. The two primary areas are shared metadata and data quality.

- **Shared metadata**—Sharing metadata between your Business Intelligence applications and Data Integrator enables you to determine the source of data in your reports and documents. Business Objects continues to enhance shared metadata to provide end-to-end impact analysis and data lineage from reports to data source. Specifically, this release enables [End-to-end metadata viewing from Desktop Intelligence documents](#).
- **Data quality**—With operational systems frequently changing, data quality control becomes critical in your extract, transform, and load (ETL) jobs. Data Integrator provides data quality controls that act as a firewall to identify and fix errors in your data. Specifically:
 - Data Integrator 11.0.1 provided the auditing data flow feature.

- Data Integrator 11.5.0 provides a [Data profiler redesign](#).
- Data Integrator 11.5.1 provides [Data Quality dashboard metadata reports](#) and [Validation transform enhancements](#).

Case preservation for database object names

Data Integrator now preserves the case of schema object names as they exist in database catalogs. These schema object types include tables (owner names, column names, etc.), functions, and domains. The result is that Data Integrator now displays the original case for object names in Business Objects Universes, Metadata Exchange, and exported .atl and XML files.

Data Quality dashboard metadata reports

Data Quality dashboard metadata reports provide graphical depictions that let you evaluate the reliability of your target data based on the validation rules you created in your Data Integrator batch jobs. This feedback allows business users to quickly review, assess, and identify potential inconsistencies or errors in source data.

After establishing validation rules in your data flows, you build custom reports by defining *functional areas* (for example “Employees”) and *business rules* (for example “Address format”). You then associate the existing validation rules with the business rules and functional areas.

Then business users, such as a Human Resources manager, can view the reports to quickly evaluate the integrity of, for example, address information in the source. They can also drill in to reports to identify specific validation rules and view sample data.

For more information, see [Chapter 6, “Data Quality Dashboard Reports,”](#) in the *Data Integrator Metadata Reports User's Guide*.

Data profiler redesign

This version of Data Integrator provides a Data Profiler that obtains information that you can use to determine:

- The *quality* of your source data before you extract it so that you can perform data cleansing or other transformations.
- The *structure* of your source data to better design your Data Integrator jobs and data flows, as well as your target data warehouse.
- The *content* of your source and target data so that you can verify that your data extraction job returns the results you expect.

The Data Profiler uses Data Integrator engine processes to execute profiler tasks that can scale in an enterprise environment. The Data Profiler tasks generate and collect the following information that multiple users can view:

- Column analysis—The Data Profiler provides two types of column profiles:
 - Basic profiling—This information includes minimum value, maximum value, average value, minimum string length, and maximum string length.
 - Detailed profiling—Detailed column analysis includes distinct count, distinct percent, median, median string length, pattern count, and pattern percent.
- Relationship analysis—This information identifies data mismatches between any two columns for which you define a relationship including columns with an existing primary key and foreign key relationship. You can save two levels of data:
 - Save the data only in the columns that you select for the relationship.
 - Save the values in all columns in each row.

The Data Profiler is part of Data Integrator, and it does not require a separate license. To use the Data Profiler, you must define a profiler repository and associate it to a Job Server and the Administrator.

For details, see [Chapter 9, “Profile Server Management,” in the *Data Integrator Administrator Guide*](#). For information about executing profiler tasks and viewing the profile results, see [“Using the Data Profiler” on page 319 of the *Data Integrator Designer Guide*](#).

End-to-end metadata viewing from Desktop Intelligence documents

This version of Data Integrator extends integrated metadata support to the entire set of Business Intelligence products that Business Objects offers.

- Data Integrator 11.0.0 extended metadata reports to include Crystal Reports.
- Data Integrator 11.0.1 further extended metadata reports to include Business Views and Web Intelligence documents.
- Data Integrator 11.5 extends metadata reports to include Desktop Intelligence documents.

This complete metadata integration allows impact analysis and lineage analysis for all report types and documents that you can create with Business Objects Business Intelligence products. Specific benefits include:

- Allows designers to understand Business Views, Business Elements, and Business Field lineage to determine which data sources a Central Management Server uses to produce a Business View.
- Lists the original source of the data if any Data Integrator data flows update the tables used by Business Elements, Business Objects documents, or Web Intelligence documents.
- If changes occur in the original source tables and columns, you can analyze dependencies to:
 - Determine which Business Views and dependent Crystal Reports are affected.
 - Determine which Web Intelligence and related Crystal Reports are affected.
 - Determine which Desktop Intelligence documents are affected.

For information about setting up the metadata integrator for Desktop Intelligence documents, see the [Data Integrator Getting Started Guide](#). For information about metadata for Desktop Intelligence documents, see the [Data Integrator Metadata Reports User's Guide](#).

Validation transform enhancements

In conjunction with the new Data Quality dashboard metadata reports, the validation transform now includes the following features.

- Validation transform options—These options enable data-quality statistics and/or sample-data collection for Data Quality dashboards.
- Validation rule properties—When defining a validation rule, you can now create a name for the rule (instead of using the automatically assigned column name) and optionally add a description. More descriptive names are useful when creating your business rules for Data Quality dashboards.

In addition, you can disable data-quality statistics collection at the job level with new options on the job execution properties pages in the Designer and the Administrator.

For more information about Data Quality dashboards, see [Chapter 6, “Data Quality Dashboard Reports,”](#) in the [Data Integrator Metadata Reports User's Guide](#).

Variable-length character processing enhancements

This version of Data Integrator provides the following changes to conform to the ANSI SQL-92 varchar behavior:

- Treats an empty string as a zero length varchar value (instead of NULL).
- When you use the operators Equal (=) and Not Equal (<>) to compare to a NULL constant, the comparison always evaluates to FALSE. Uses new IS NULL and IS NOT NULL operators in Data Integrator scripting language to test for NULL values.
- Treats trailing blanks as regular characters, instead of trimming them, when reading from all sources.
- Ignores trailing blanks in comparisons in transforms (Query and Table_Comparison) and functions (decode, ifthenelse, lookup, lookup_ext, and lookup_seq).

For more details, refer to [“varchar” on page 238 of the Data Integrator Reference Guide](#).

If you are upgrading from a previous version of Data Integrator, the default is to use the existing varchar behavior for backward compatibility. However, Business Objects recommends that you use the ANSI varchar behavior because the previous varchar behavior will not be supported in a future Data Integrator version. For details, see [“Migration considerations” on page 23 of the Data Integrator Release Notes](#).

Productivity

The following products and features can enhance your productivity when working with Data Integrator.

BusinessObjects™ Composer

BusinessObjects™ Composer is a new, independent application that can significantly improve the way you design and develop ETL models for your data-warehousing projects. Developing higher-level designs for your ETL projects makes the process more efficient and improves the quality of resulting implementations. Composer complements ETL tools by providing a platform for creating a design much earlier in the process of building your data-warehousing project. It allows you to identify and better understand sources of data, implement transformations, and document your designs.

Composer’s intuitive, web-based user interface promotes graphically building end-to-end data warehousing projects. Additional features include:

- **Profiling**—The Data Profiler helps you determine the quality of your data. For more information, see [“Data profiler redesign” on page 7](#).
Note: You must have profiling configured in Data Integrator to access it in Composer.
- **Browsing/Exploring**—Composer allows you to browse and explore metadata in your databases and other ERP sources to enable you to better understand your technical and business metadata including datastore system views and a project explorer.
- **Reporting**—You can generate reports on a system-wide basis or in a project-specific basis. Report types include project overviews (for one or all projects) or reports with compiled user, task list, or document Information.
- **Subject areas**—To better manage your sources, you can create a subject area. A subject area is a collection of tables that come from one or more data sources that apply to a specific subject. Use subject areas to group tables into logical units.
- **Integration with Data Integrator**—Composer is complementary to Data Integrator. After creating a project and developing your mappings, you can generate a “starter” job for use in Data Integrator, thereby even further reducing the time required to implement your project.
- **Exporting and importing**—You can export or import the contents of a Composer project via an XML-based format, which allows you to move your projects and definitions to other systems.

COBOL copybook file format enhancements

Data Integrator supports the widely used COBOL copybooks as flat file sources. The COBOL copybook file reader now supports two new features:

- **Field ID**— If you have multiple record types in one file (for multiple 01-level record definitions), the Field ID option allows you to create rules for indentifying which records represent which schemas.
- **Record Length Field**—This option lets you identify the field that contains the length of the schema’s record.

For details, see [“Creating COBOL copybook file formats” on page 145 of the *Data Integrator Designer Guide*](#) “ and [“COBOL copybook file format” on page 40 of the *Data Integrator Reference Guide*](#).

Function enhancement (rand_ext)

Data Integrator now contains a more powerful function for generating random number results.

For details, see “[rand_ext](#)” on page 516 of the *Data Integrator Reference Guide*.

Metadata reports redesign

Metadata reporting capabilities have been completely redesigned and expanded in Data Integrator.

The modules new to Data Integrator version 11.5.0 are:

- Impact and Lineage
- Operational Dashboards
- Auto Documentation

The module new to Data Integrator version 11.5.1 is:

- Data Quality dashboards

For more information about metadata reports, see the *Data Integrator Metadata Reports User's Guide*.

Impact and Lineage

The redesigned Impact and Lineage module of metadata reporting provides an intuitive, easy-to-use hierarchical interface for viewing your Data Integrator datastores, Business Objects CMS servers, and associated components.

Select an object in the left pane to display summary, impact, or lineage information for the object in the right pane.

For CMS servers, you can configure and run the Metadata Integrator to view metadata reports for Business Views, Crystal Reports, Universes, Desktop Intelligence documents, and Web Intelligence documents.

Operational Dashboards

The new Operational Dashboards module of metadata reporting provides graphical and tabular details and summaries about job execution across your projects. These reports help you manage your job execution to improve efficiency and performance. The Operational Dashboards home page displays job execution statistics and duration in both a current time frame (like a snapshot) or over a longer period (for trend analysis). You can drill in to these reports for more detailed reports, many of which display in both graphical and tabular formats.

Data Quality dashboards

This module is new to Data Integrator XI Release 2 version 11.5.1. Data Quality dashboard reports provide graphical depictions that let you evaluate the reliability of your target data based on the validation rules you created in your Data Integrator batch jobs. This feedback allows business users to quickly review, assess, and identify potential inconsistencies or errors in source data. See also [Data Quality dashboard metadata reports](#) under [Trusted information](#).

Auto Documentation

The goal of the new Auto Documentation module is to provide a means to document (for example, print) representations of your Data Integrator projects. The Auto Documentation page hierarchically displays projects, jobs, work flows, and data flows as in Data Integrator Designer. For example, you can expand or collapse these entities or drill into the diagrams to customize a representation.

Query transform enhancements

When you change an input schema to the Query transform, Data Integrator now checks the existing top-level mappings to determine if any remapping is required. If the mapping contains a column with a table name that is not a current input schema name and the column is in the new input schema, Data Integrator automatically replaces the table name with the new input schema name. For some cases where Data Integrator does not automatically remap, you can use a new **Schema Remapping** option.

In addition, the **From**, **Group By**, and **Order By** tabs now provide options to change the order of tables and columns that you specify on those tabs.

For more information, see [“Query” on page 327 of the Data Integrator Reference Guide](#).

Tutorial upgrade

This version of Data Integrator expands the exercises in the [Data Integrator Core Tutorial](#) to include new material that covers the following features:

- A new Data Quality chapter that introduces a subset of Data Integrator features you can use to verify and improve the quality of your data:
 - Profile data
 - Validation transform
 - Auditing a data flow feature

- Audit details report in the Operational Dashboard module of metadata reports
- Impact and Lineage module of metadata reports
- XML_Pipeline transform

Several exercises were updated to provide more clarity (for example, the `lookup_ext` function).

After installing Data Integrator, the Core Tutorial is available from the Data Integrator Designer Help menu, the Doc\Books directory in your Data Integrator installation, or from the Windows Start > Programs menu.

Web Services option enhancement

A new option in the Administrator's Web Services Configuration page enables access to full batch job attributes. When selected, the new **Enable full batch job attributes** option allows the input message for all the batch jobs you publish to include all of the options supported for submitting batch jobs from the Administrator.

For details, see ["To configure Web service information using the Administrator"](#) on page 152 of the *Data Integrator Administrator Guide*.

Windows clustering failover support

Data Integrator Services can now utilize failover support in a Windows Clustering Environment. In the event of a hardware failure or Windows software failure, the Windows Cluster Manager will attempt to restart your Data Integrator Services.

After you create a Windows cluster, simply install Data Integrator on a shared drive from the first cluster computer, create a new resource for the Data Integrator Web Server Service as a Generic Service, then run the Data Integrator cluster installation utility to populate the other cluster nodes with the Data Integrator Service-related information.

For more information, see ["Create a Windows cluster \(optional\)"](#) on page 53 of the *Data Integrator Getting Started Guide*.

Scalability

The following features can improve the scalability of your Data Integrator projects.

Changed-data capture (CDC) enhancement

RDBMS vendors are adding more features to their databases that allow third-party applications like Data Integrator to manage the CDC environment. This version of Data Integrator extends the native CDC feature to the asynchronous publishing modes of Oracle 10g.

- Data Integrator 6.5.0 introduced the native CDC feature for Oracle synchronous publishing mode.
- Data Integrator 11.0.0 extended the native CDC feature to include Microsoft SQL Server, IBM DB2, and real-time mainframe CDC via third-party partnerships.
- Data Integrator 11.5.1 further extends the native CDC feature to include the asynchronous publishing modes of Oracle 10g. The asynchronous modes capture the changed data offline, which improves performance over synchronous mode on the source database.

For details, see [Chapter 19, “Techniques for Capturing Changed Data,”](#) in the *Data Integrator Designer Guide*.

Oracle Real Application Cluster (RAC) support

This version of Data Integrator Administrator now supports connections to Oracle Real Application Clusters (RAC). In an Oracle RAC system, multiple Oracle instances (that are running on different CPUs) access a single physical Oracle database.

An Oracle RAC system provides server load balancing. For example, if network traffic suddenly increases, Oracle RAC can distribute the load over many CPUs. You can use connection failover and client load balancing with an Oracle RAC system, but they are not part of Oracle RAC.

For more information, see [“Connecting repositories to the Administrator”](#) on page 24 of the *Data Integrator Administrator Guide*.

Performance improvements

Data Integrator writes statistics about work flows, data flows, and transforms into the AL_STATISTICS table. Now, Data Integrator writes those statistics in constant time (under one minute) regardless of the AL_STATISTICS table size or complexity of your job. Therefore, job execution performance is significantly faster.

Also, Data Integrator now provides a way to control the size of your AL_STATISTICS table by disabling transform statistics collection and collecting only work flow and data flow statistics. To use this option, in the Designer go to the Tools menu and select **Options > Job Server > General**. Then enter the appropriate value for each parameter as in the following table:

Parameter	Value
Section	AL_Engine
Key	Disable_Transform_Statistics
Value	TRUE

Teradata throughput improvement

Data Integrator now supports bulk loading with named pipes using the Teradata Warehouse Builder and the Teradata load utilities.

For details, see “Bulk loading in Teradata” on page 56 of the *Data Integrator Performance Optimization Guide*.

XML_Pipeline transform

With the new XML_Pipeline transform, you can now effectively process large amounts of XML. This simple transform processes a small portion of the XML input at a time, constantly freeing up memory to keep your XML data processing through the data flow.

For more information on this new transform, see “XML_Pipeline” on page 374 of the *Data Integrator Reference Guide*.

