Retek® Extract Transform and Load 10.3.3

Release Notes

Retek

Corporate Headquarters:

Retek Inc.
Retek on the Mall
950 Nicollet Mall
Minneapolis, MN 55403

888.61.RETEK (toll free US)
+1 612 587 5000

European Headquarters:

Retek
110 Wigmore Street
London
W1U 3RW
United Kingdom

Switchboard:
+44 (0)20 7563 4600

Sales Enquiries:
+44 (0)20 7563 46 46
Fax: +44 (0)20 7563 46 10

Customer Support

Customer Support hours:

Customer Support is available 7x24x365 via e-mail, phone, and Web access.

Depending on the Support option chosen by a particular client (Standard, Plus, or Premium), the times that certain services are delivered may be restricted. Severity 1 (Critical) issues are addressed on a 7x24 basis and receive continuous attention until resolved, for all clients on active maintenance.

Contact Method	Contact Information
Internet (ROCS)	www.retek.com/support Retek's secure client Web site to update and view issues
E-mail	support@retek.com
Phone	US & Canada: 1-800-61-RETEK (1-800-617-3835) World: +1 612-587-5800 EMEA: 011 44 1223 703 444 Asia Pacific: 61 425 792 927
Mail	Retek Customer Support Retek on the Mall 950 Nicollet Mall Minneapolis, MN 55403

When contacting Customer Support, please provide:

- Product version and program/module name.
- Functional and technical description of the problem (include business impact).
- Detailed step by step instructions to recreate.
- Exact error message received.
- Screen shots of each step you take.

Contents

RETL 10.3.3 Overview

This release includes several critical bug fixes.

The first bug addresses a buffer overflow problem in oraread. If the SQL select statement in the 'query' property of oraread exceeds a certain length, RETL will abort with a core dump. The fix allows a 'query' property of arbitrary length.

The second bug fix corrects an internationalization problem where RETL incorrectly parsed an EOF when reading certain regional character codes.

The third bug fix corrected a memory corruption problem when RETL would read certain character codes in a flow file. This behavior was noted in the oraread operator's 'query' property as well as in generator's 'schema' property.

Compatibility Matrix

The following represents OS/Database combinations that are supported and have been certified to work properly :

OS	Version	Arch	DB	Version	Executable
AIX	4.3.3	64	Oracle	9.2	rfx.a433.o9x.64
AIX	4.3.3	64	Oracle	9.0.1	rfx.a433.o9x.64
AIX	4.3.3	32	DB2	7.2.3	rfx.a433.d723.32
AIX	4.3.3	32	Oracle	8.1.7	rfx.a433.o817.32
AIX	4.3.3	32	Tera	2r4.2	rfx.a433.t241.32
AIX	4.3.3	32	None	None	rfx.a433.nodb.32
AIX	4.3.3	64	None	None	rfx.a433.nodb.64
AIX	5.1	64	Oracle	9.2	rfx.a51.o9x.64
HP-UX	11i	64	None	None	rfx.h11i.nodb.64
HP-UX	11i	32	None	None	rfx.h11i.nodb.32
HP-UX	11i	64	Oracle	9.2	rfx.h11i.o9x.64
HP-UX	11i	64	Oracle	9.0.1	rfx.h11i.o9x.64
HP-UX	11i	32	Oracle	8.1.7	rfx.h11i.o817.32
HP-UX	11i	32	Tera	2r4.2	rfx.h11i.t241.32
Solaris	8	32	None	None	rfx.s58.nodb.32
Solaris	8	64	None	None	rfx.s58.nodb.64
Solaris	8	32	DB2	7.2.3	rfx.s58.d723.32
Solaris	8	64	Oracle	9.2	rfx.s58.o9x.64
Solaris	8	64	Oracle	9.0.1	rfx.s58.o9x.64
Solaris	8	32	Oracle	9.0.1	rfx.s58.o9x.32

OS	Version	Arch	DB	Version	Executable
Solaris	8	32	Oracle	8.1.7	rfx.s58.o817.32
Solaris	8	32	Tera	2r4.2	rfx.s58.t241.32

Note: Any entries that have 'none' listed as the database means that the corresponding executable can run standalone. For example, if RETL is to be used standalone on Solaris 8 32bit, the 'rfx.s58.nodb.32' binary would be used.

Installation

You can get the download from "http://mspdev25:8080/retl/". This release is broken up into the builds for different platforms for convenience -- one for each of HP-UX, SunOS and AIX. These have an .h11i, .s58, and .a433 extension respectively.

You can check the contents of the package and verify that it is valid by executing the following command:

```
gunzip -c <package.tar.gz> | tar -tf -
```

To extract the package use the following command:

```
gunzip -c <package.tar.gz> | tar -xf -
```

This creates the directory and the package contents (see below).

Package Contents

The following is a brief description of the contents of retl.10.3.2.tar.gz:

bin/

Includes executables required for running retl on different platforms.

File	Description
gsort.SunOS	gsort for SunOS
rfx.s58.o817.32	32bit RETL built for Sun Solaris 8, Oracle 8.1.7
rfx.s58.o9x.32	32bit RETL built for Sun Solaris 8, Oracle 9.x
rfx.s58.o9x.64	64bit RETL built for Sun Solaris 8, Oracle 9.x
rfx.s58.t421.32	32bit RETL built for Sun Solaris 8, TeraData 2v4.1
rfx.s58.d723.32	32bit RETL built for Sun Solaris 8, DB2 v 7.2.3
gsort.HP-UX	gsort for HP-UX
rfx.h11i.o817.32	32bit RETL built for HP-UX 11i, Oracle 8.1.7
rfx.h11i.o9x.64	64bit RETL built for HP-UX 11i, Oracle 9.x
rfx.h11i.t421.32	32bit RETL built for HP-UX 11i, TeraData 2v4.1
gsort.AIX	gsort for AIX
rfx.a51.o9x.64	64bit RETL built for AIX 5.1, Oracle 9.x

File	Description
rfx.a433.o817.32	32bit RETL built for AIX 4.3.3, Oracle 8.1.7
rfx.a433.o9x.64	64bit RETL built for AIX 4.3.3, Oracle 9.x
rfx.a433.t421.32	32bit RETL built for AIX 4.3.3, TeraData 2v4.1
rfx.a433.d723.32	32bit RETL built for AIX 4.3.3, DB2 v 7.2.3
verify_retl	The RETL verification tool
README.verify_retl	The README for verify_retl

Note that the Oracle 9.x builds were tested (but not certified) on both 9.0.1 and 9.2 Oracle datbases. The AIX 4.3.3 builds were tested under AIX 5.1 and work except for the 64 bit version which is why we have a build for that version. See the compatibility matrix for more information.

lib/

These are dynamically linked libraries that are needed in order to run RETL.

File	Description
libKCC-eh-ts.sl.hpux.32.kcc	32 bit KCC library for HP-UX 11i

docs/

This directory contains the documentation associated with RETL.

File	Description
Retl-1033-rn.txt	This document file.

etc/

This directory contains the documentation associated with RETL.

File	Description
rfx.conf	The default retl configuration file.

samples/

This directory contains the samples packaged with RETL.

See the README in the samples/ directory for more information about the samples packaged with RETL. These are the samples that are run after the install, by verify_retl

Changes

Changes since 10.3.2

Fixed #398

Fixed i18n problem where RETL incorrectly recognized an EOF character in an embedded character sequence.

Fixed #399

RETL would core dump when the SQL select statement in the 'query' property of the oraread operator was too long. Fixed the buffer overflow condition to prevent core dump and allow RETL to use SQL queries of any arbitrary length.

Fixed #401

Memory corruption would cause a core dump in certain flows. The problem was related to an attempt to write past the end of an array, which caused RETL to core dump in another part of the code. The offending code was rewritten and now RETL will no longer core dump in this situation. Affected operators were: oraread, generator, convert.

Changes since 10.3.1

Fixed #387

Added the new properties in orawrite operator to allow users to utilize Oracle SQLLoader's delimiter format for better performance.

Fixed #388

Added a property in orawrite to stop SQLLoader process if the number of bad data happens.

Changes since 10.3

Fixed #180

Oraread couldn't handle long queries or many cols in descriptors. Fixed a dangerous buffer overflow condition that could occur when oraread's select statement was over a certain length.

Fixed #379

RETL would core dump when the create table statement was too long in orawrite operator. Fixed a buffer overflow condition that occurred when the create table statement of the orawrite operator became too large.

Changes since 10.2.1

Added logging to RETL. RETL will now print start and stop times as well as when specific operators start, finish, and how many records they process depending upon specified logging options which can be controlled via the config file. This closes issues #141 and #176.

Fixed #179

Added an option "-sNONE" to prevent schema information from being output upon initialization. Additionally, a "-sSCHEMAFILE" option has been added to allow speedy production of schema files.

Fixed #181 and #337

fieldmod rename doesn't error when new column is nonexistent. A better error has been introduced to fieldmod that catches a variety of errors that were previously not checked for.

Fixed #243

When you type "rfx -h" it now shows correct and working default values.

Fixed #251

The convert operator will now be able to convert uint64 to other numerical data types.

Added many default conversions

Added default conversion to uint64 from each of: int8, uint8, int16, uint16, int32, uint32, int64, sfloat, and dfloat. Added default conversion from each of: int8, uint8, int16, uint16, int32, uint32, int64, uint64, sfloat to dfloat. Added default conversion to int64 from each of: int8, uint8, int16, uint16, int32, uint32, uint64, sfloat, and dfloat.

Closed #258

Added "PARTITIONING KEY" on db2write when creating tables. This parallelizes the data across nodes to increase performance. This improves performance upon table creation.

Closed #275

Validate unknown fields in schema files.

Closed #293

All dbwrite operators now write to a backup file when RFX_SHOW_SQL and RFX_DEBUG are set. Also added additional log messages when RFX_SHOW_SQL=1 and RFX_DEBUG=1 to help user recreate the sql load.

Fixed #310 and #352

RETL now does explicit type checking to ensure that data types within the export schemas match the actual outgoing data. If this is not the case, an error message will be displayed and RETL will terminate abnormally. Previously if types differed between input and output schemas, RFX would output corrupt data or core dump.

Fixed #313

In TeraWriteOperator, we now only write NULLIF='' if a field is nullable, rather than writing NULLIF='<max field length>'. This solves the problem with NULLIF hitting the maximum length of 80 characters on fields longer than this.

Fixed #316

The keep property is now case-insensitive in the removedup operator.

Fixed #320

RETL for Oracle 64-bit now shows "Flow ran successfully" when a flow completes successfully.

Closed #324

Oraread 'query' property can now read unicode strings. This is related to an enhancement that allows RETL flows to be specified via UTF-8.

Closed #325

RETL now supports multi-byte characters via the "bytes_per_char" property in the configuration file. See the Programmer's Guide for more details.

Closed #327

In previous releases, rfx was returning 0 when it failed to create threads for large flows on certain platforms with multiple partitions. RETL will now return an error if it fails to create a thread.

Fixed #331

RETL now limits the uses of memory and pages data out to disk to ensure that a limit is placed upon RAM consumption. Previously it was possible (though rare) that RETL's memory consumption could grow to consume all available memory.

Closed #332

A sensible error message is now displayed if the output file property of an export is misspelled, empty, or forgotten.

Closed #335 and #340

Cliprows no longer allows negative values in cliprows operator.

Fixed #342

RETL does additional checking on schema files. Specifically, 'name' must be specified for each field in the schema, 'delimiter' must be specified for each delimited field in the schema, and 'nullvalue' must be specified for each field that has a 'nullable' attribute. Appropriate error messages will be displayed if these conditions are not true.

Changes since 10.2

Fixed #331

- RETL was consuming large amounts of memory and in some cases crashing the server. This was a result of a circular dependency which caused dataset expansion when it did not need to occur. This problem has been resolved.
- Allocates memory chunks more conservatively. This results in significantly better use of memory. In many cases, flows run with 30-50% less memory than in release 10.2 and 1.7.1.
- Adds the ability to clean up memory as operators finish. This reduces overall memory consumption and allows memory to be released prior to completion of the entire flow.
- Improved efficiency of locking within Datasets. This improves overall efficiency of data processing and results in a 10-15% performance increase. This change should also allow better scale because operators no longer need to synchronize on a per record basis. In short, the more complex the flow, the more noticeable the improvement should be.
- A fix to the AIX 4.3.3 platform allows the use of an enhanced DataSet. In 1.7.1 and 10.2 releases a timed mutex wait was required within the more advanced datasets. This operation caused a significant degradation in performance on AIX 4.3.3. As a result, RETL used a less efficient DataSet for this platform. This problem has been fixed by removing the need for the timed wait. This should result in significant performance improvements for that platform.

Changes since 1.7.1

- Oraread operator will be able to run stored procedures if "sp_prequery" and /or "sp_postquery" properties are specified. One of its usages is flashback for Oracle 9i. See the Programmer's Guide for more details.
- Several performance enhancements that should result in a 100-400% speed increase over previous versions. That is, rfx should run between 2-5 times faster. Generally this is more noticeable on more complex flows with larger numbers of records.
- The new property "primaryindex" has been added to the terawrite operator to let users specify primary index for the created table. This will improve performance when they create or recreate tables. If the users don't specify the "primaryindex" property, the first column in the table will be used as the primary key.

TERAWRITE

- Will now use fastload rather than mload when writing to empty tables. This should result in better performance in that specific case.
- Improved messages within TERAWRITE when RFX_SHOW_SQL is exported - shows when MLOAD and FASTLOAD are being called.
- If RFX_DEBUG and RFX_SHOW_SQL are specified, TERAWRITE will write the data that is being written to the DB also to a flat file for debugging.
- Improved rfx so that dynamic library dependencies are minimized. This should reduce/eliminate many of the dependencies on environment variables.
- There must be a 1-to-1 correlation between INPUT and OUTPUT datasets. Exceptions to this rule are now caught by RETL and reported to the user.
- Support for Oracle 9.2 -- see the compatibility matrix for more information.
- Support for AIX 5.1L -- see the compatibility matrix for more information.

Fix for bug #186, #271

- Differences in int64 on 32bit and 64bit platforms. Fixed major problems with int64/int32 in that underflow/overflow conditions weren't being checked. This resulted in garbage data being returned in conversions of large numbers. This bug is closely related to bug #271 on HPUX.
- DB2Write now defaults to conventional mode loading. It is safer; only those who really need to optimize should be worrying about direct mode loading.
- Better error messages when sql loader fails. Includes more information (logfile and control file, rather than just parfile) when RFX_SHOW_SQL=1
- rfx no longer accepts blank lines as valid input.
- Allow "" to mean a blank string for nullvalue within fixed length records (within schema files). Previously, rfx required the developer to specify a nullvalue of exactly the same length as the fixed length field. Thus for a 500 character field, rfx previously required that RETL developers specify: "<insert 500 blanks here>". Developers may now simply enter "" to mean blanks for the nullvalue with the number of blanks equivalent to the length of the fixed length field.
- Modified filter operator to alert user when multiple outputs have been specified but 'rejects' property is not specified or set to false. Previously the user would have seen the message: 'Num outputs does not match attached datasets'

Fixed #202

Better detection and reporting of field lengths and other properties.

Fixed #222

When converting from a number to a string, the string did not receive a maxlength which is necessary for writing to a database. The work-around for this was to have developers to break flows into multiple flows which have a big impact on the overall efficiency of the batch process. Addition of this fix should allow developers to collapse flows into one flow that is capable of converting and writing to the database.

Fixed #235

Permissions of any files RETL uses are now checked. This includes making sure flows/schemas/configuration files/import files are readable, and any temp space RETL uses is writable. Appropriate error messages are displayed if files don't have the correct permissions.

Fixed #246

During an import, if a null value existed in a date field and the field was not nullable a seg fault/core dump occurred. Now the record is rejected since it does not meet the schema specification.

Fixed #247

This caused multiple string statements within Filter causing RFX to fail with this message: "Invalid String passed to filter. String must be enclosed in single quotes." Filter has been corrected to handle the problem without errors.

Fixed #249

Import operator now generates an error message when a new line is the last character of the import data file. This empty line is equivalent to an empty record and is rejected as not meeting the file schema.

Fixed #256

Filter now treats all null values as being less than numeric values. Previously, it would have treated 100 (for example) as being less than a null value.

Fixed #262

If the "filter" property of a FILTER operator is empty, the following error will occur "RFX Unexpected Error: FilterOperator.cpp:119: '[filter:0]: Empty value for 'filter' property.'"

Fixed #270

Some error messages were being broken up when printed.

Fixed #290

Corrected a memory problem that occurred when delimited non-string fields were too long (invalid data). The result was a garbled error message.

Fixed #304

Environment checks to make sure RFX_HOME, ORACLE_HOME, and other variables needed are set in order to run RETL properly. The user is now given better error messages when these variables aren't set.

Fixed #222

The maxlength of a string is set when a non-string is converting to a string in convert operator.

Fixed #296

We isolated and fixed a very low probability condition that occurred much more frequently when bufsize was set to a very low number. There are no longer deadlocks even when bufsize is set to a very small number.

Fixed #277

Time and Timestamp fields are now validated and will be dealt with as improperly formatted fields upon IMPORT.

Changes since 1.7.0

Fixed bug #241

Allows a flow that failed during TeraWrite operator load to be rerun without manually dropping tables and releasing mload.

Fixed bug #238

Deadlock was occurring when the bufsize was reduced to a very small number. In short, a signal to expand datasets was not being caught by the dataset and therefore datasets were not being properly expanded. This is now resolved.

Fixed bug #230

Previously there could be a name collision in the temp files when multiple versions of rfx were run simultaneously. The current version ensures that this is not possible by using PID and thread# when creating file names.

Fixed bug #123

Large floating point numbers were not being accurately output in fixed length fields. Additionally, large floating point numbers were not being accurately represented.

Floating point numbers will now use regular formatting until that formatting would exceed the length of the fixed length field, at which point rfx will use scientific notation dropping least significant digits if necessary to accommodate the field length.

Fixed bug #225

UserNames and passwords were being passed to SQL*Loader on the command line. These were viewable by doing a "ps-aef" - making this a security problem. This release moves the userid and password into a file. Assuming that permissions are set up properly for the RETL user, this resolves the problem.

Fixed bug #242

- The collect operator no longer causes deadlock. Collect and Funnel are now aliases of each other; they do exactly the same thing.
- The diff operator is now useable. It was previously unrecognized in rfx because it was not registered at initialization time.
- Tested against Oracle 9.2 Database.
- Works fine as long as executable is linked against proper version of dynamic link library. For example, retl.s58.o901.64 must be linked against the 64 bit version of the Oracle 9.0.1 libraries. Linking to the incorrect library will cause a core dump upon startup.
- Added the "samples" directory with baselines to show expected output so that developers have example flows to start from and compare against.

Changes since 1.6.3

Fixed bug #174

RFX will now exit with an error message if all INPUT datasets are not matched to output datasets.

Fixed bug #195

When converting int from string, didn't deal with zero length integer fields properly. Converted them to zero value and it now treats them as errors unless an empty string is nullvalue.

Fixed bug #203

- Blank lines in input files are no longer interpreted as records; instead they are rejected.
- Bug in Field conversion - null string was always interpreted as null value.
- Fixed bug in record parsing - unable to detect null value in last field of delimited record if empty.
- Errors are now printed to std error instead of std out.
- Fixes operators; groupby, covert, cliprows, diff, removedup, merge, lookup, and hash in rare instances where null fields were not being treated properly.

Fixed Bug #191

The DataSet expansion algorithm caused corruption in dataflows. This occurred only when immanent deadlock is detected --triggering DataSet expansion in order to keep the flow moving. It occurred in a diamond flow which is a flow that is broken apart into parallel transformations and then reassembled sometimes resulting in a circular dependency between producers and consumers.

Fixed Bug #192

Deadlock occurred with some flows when partitioning was enabled. The issue was fixed by using the funnel operator as the default collector instead of the collect operator. The problem with the collect operator still exists and use of this operator has been deprecated (use funnel instead).

Fixed Bug #193

- If the date field is nullable and the incoming data is null then "invalid date" was displayed in the result field. This needs to be handled differently for fixed length fields.
- Correction of a DB2 issue that resulted in dead lock while trying to drop and recreate tables; DB2 table creation has been serialized so that creation is attempted only once.
- rfx now works properly with TeraData on HP-UX.
- Orawrite will report an sql loader error when a rebuild index fails during a direct load.
- The import operator now rejects records that have characters in a numeric field.
- Better error reporting of exceptional conditions.
- SQL statements to and from databases are turned off by default unless an error or abnormal condition occurs. These can be turned on for debugging purposes by exporting RFX_SHOW_SQL on the command line.
- Developers can observe DataSet FULL, EMPTY, NOT FULL, NOT EMPTY and EXPANSION events by exporting RFX_DEBUG_DEADLOCK.

Changes between 1.6.2 and 1.6.1

- If all the rows in a column are null, the GroupBy operator will provide a null column total.
- An added property that allows export operator to append data to an existing data file.
- An operator was added that is similar to the GroupBy called 'ClipRows', allowing the last or first N records to be kept.
- When using an import operator for a fixed length file, if the data of dfloat field is right justified, the data will be processed.
- The import operator rejects records whose data length exceeds the schema definition.
- RETL returns an error code if no data file or schema file found in the import operator. Also includes improved error message including file name and operator name.
- The import operator now includes an option to strip leading and/or trailing white space from string columns.
- A syntax error is generated if users don't specify the maxlength for the string in import operator.
- A error code of '1' (previously '0') is returned when users give a wrong path or file in import and export operators. Additionally, whenever there is an error condition RETL will return a non-zero value.

- The sort operator was corrected to sort in the correct order for string types.
- RETL error handling messages have been improved in many different areas to include more descriptive text. This is an ongoing improvement.

Known Issues

- Sortfunnel may not maintain sorted order when input values are null. This has been known to affect any flows that try to sortfunnel records that contain null int fields. The workaround is to collect and then sort instead of using sortfunnel when this problem is encountered.
- Translation from number to string for large arbitrary precision floating point numbers.
- When specifying fields in filter operator, they must all be uppercase or they will not be recognized.
- The dropedit of the diff operator does not work properly.

Bug #171

RFX leaves some temporary files in tmp directories. These directories should be purged on a regular basis until this issue is resolved.

Bug #182

- Properties are not properly validated.
- There were more than 60% of property names and values that were passed to RETL without basic error checking. This means current RETL applications may not function properly or may output wrong data. It is hard debug. Here are a few examples:

1 If you set up a Boolean property value to "true" or "TRUE", RETL will have true for the value; otherwise it will have false value no matter what. If you spell "TRUe", RETL will get false.

2 If you spell a wrong value for mode property in database operators, RETL will use a default value ("append" in orawrite and "insert" in db2write). The same is true for method and createtable properties in database operators and in many cases in other operators

3 If you set up the "desc" value for order property in sort operator, RETL will sort by desc order; otherwise it will sort by asc order no matter what. If you type "dess", it will sort by asc. There is the same problem in sortfunnel.

4 If you don't spell the property names correctly, RETL may use default or has no function for the properties without any error message.

5 If you specify a removedup property in sort operator no matter what value you set up, RETL will treat it as true and remove duplicated record. It does what it is said in Programmer's Guide, but it should allow users to change it to false.

6 Core dump or get rfx unexpected errors. In the RETL 10.3 release, we checked all property names and possible values to make sure they match what is in the Programmer's Guide and generate WARNING messages if they do not match. We added many other error checking and warning messages for other cases in this release. We recommended that you fix all WARNINGS given by RETL 10.3. Future versions of RETL may trigger an error it these conditions are not fixed.