IRI FACT
Fast Extract for DBs

Product Overview

Technical Summary, Samples, and Specifications

IRI
Total Data Management

Introduction

IRI Fast Extract (FACT) Version 3.7 is a high-performance unload utility for very large databases (VLDBs). FACT is primarily for data warehouse ETL architects and database administrators interested in bulk database migration, replication, archive, or offline reorgs.

FACT unloads table data into flat files using native connection protocols and parallel query techniques. It simultaneously creates metadata of the extracts for data transformation and bulk (re)loading.

FACT provides high-speed extraction for the following VLDB systems:

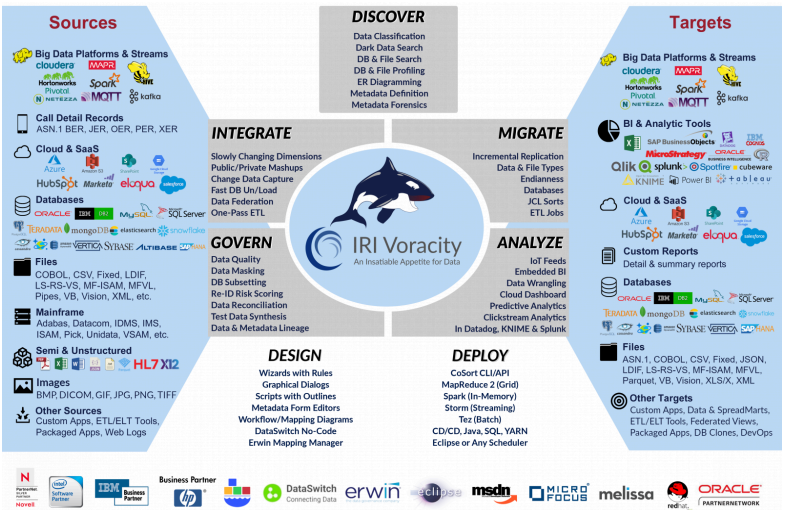
Oracle	DB2 UDB	MS SQL	Greenplum
Sybase	MySQL	Altibase	Teradata

On the front-end, FACT uses simple configuration scripts to define the database connection, parallel unload method, formatting statements, and a query in SQL SELECT syntax. The scripts can be written and modified in a text editor, or automatically in Eclipse using IRI Workbench.

On the back-end, the FACT executable consumes only a small amount of system resources, without the need for load-balancing. It accelerates extraction of table contents using multiple query splitting methods, and all supported combinations of database-native features, such as:

- parallel extraction of a partitioned table
- extraction using indices
- conditional extraction
- high-speed fetching using hints.

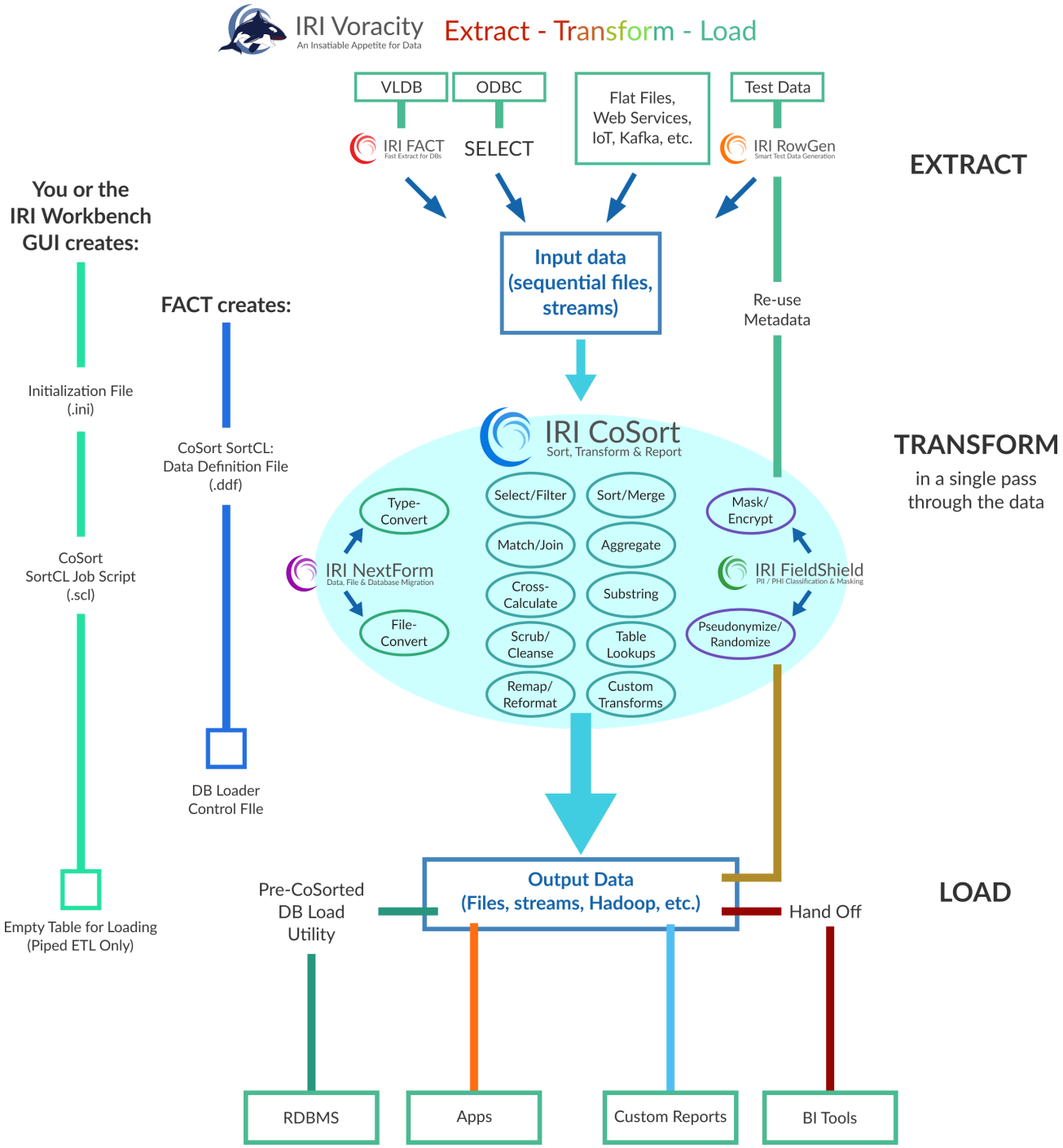
FACT also creates SortCL data definition file (.DDF) metadata for IRI CoSort and its spinoffs (NextForm, FieldShield, and RowGen), as well as DB load config files. FACT is also an optional but supported component in the IRI Voracity ETL and data management platform.



BI/DW architects and DBAs can use the SortCL metadata in Voracity fast ETL, DB migration, and offline reorg jobs. Scheduled and piped executions can run in batch or from IRI Workbench.

Unlike other database unload methods (e.g., Oracle data pump), FACT creates *portable* flat files. Its 'dump-table-to-file' data is thus readily available in or outside of IRI Voracity facilities to speed and perform: offline reorgs, transforms, pre-load sorting, migrations, change and summary reporting, ETL, replication, testing, and protection.

If you also use the IRI Voracity platform or its embedded CoSort product, the SortCL program they share will perform or accelerate all the post-extraction steps at once. But you do not have to use SortCL; once the data is in flat files, you can do anything you want with them.



You can also use FACT as a remote service by installing a Secure Shell (SSH) or daemon (SSHD) service, or a remote systems explorer (RSE) connection to your server from IRI Workbench. By connecting to another server, and specifying those details in the FACT configuration file, you can extract data from remote database tables.

Using FACT

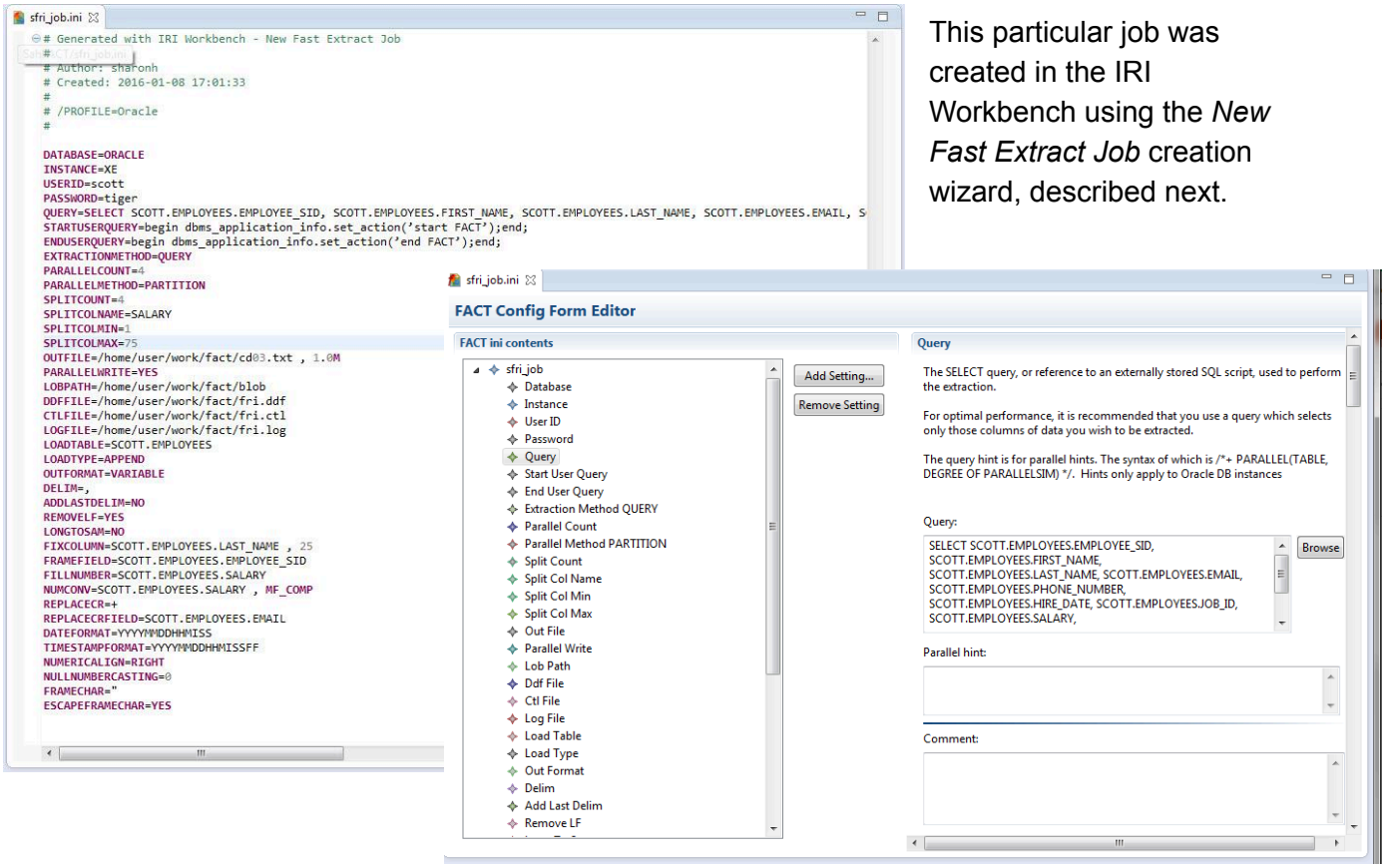
FACT is a command-line executable that relies on a self-documenting configuration file in .ini or .xml format where you specify the:

1. connection parameters to the database and table
2. query in SQL SELECT syntax and parallel unload logic
3. output formatting options
4. target file (data, runtime log, DDF and loader metadata, and LOB paths) assignments

You can create these files in any text editor, or in the free IRI Workbench graphical user interface (GUI), built on Eclipse™, using any of these components:

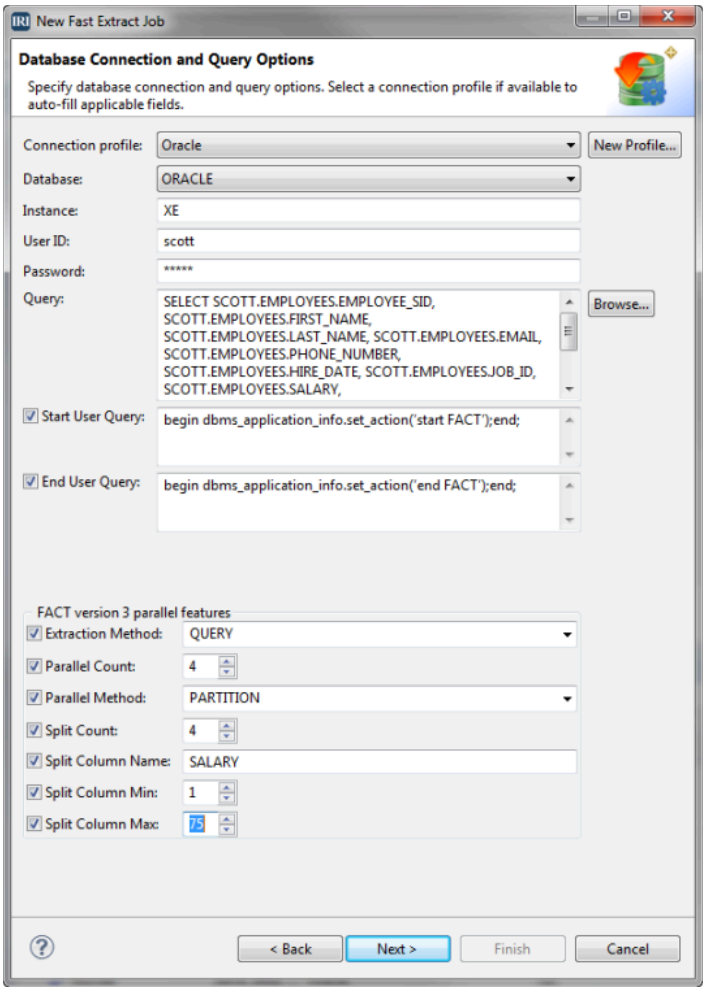
- Syntax-aware text editor
- FACT Config Form editor
- *New Fast Extract job* … wizard
- *New Multi-Table Reorg Job* … wizard

Following is an example of a FACT .ini file, open in the IRI Workbench script editor for FACT, with colored keyword and syntax validation:



This particular job was created in the IRI Workbench using the *New Fast Extract Job* creation wizard, described next.

FACT Job Creation Wizard



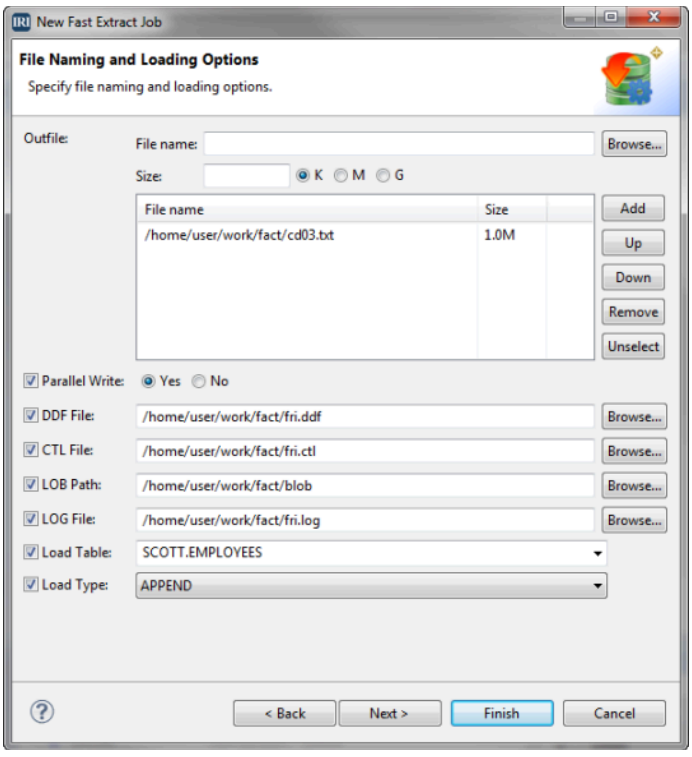
Connection and Query Options

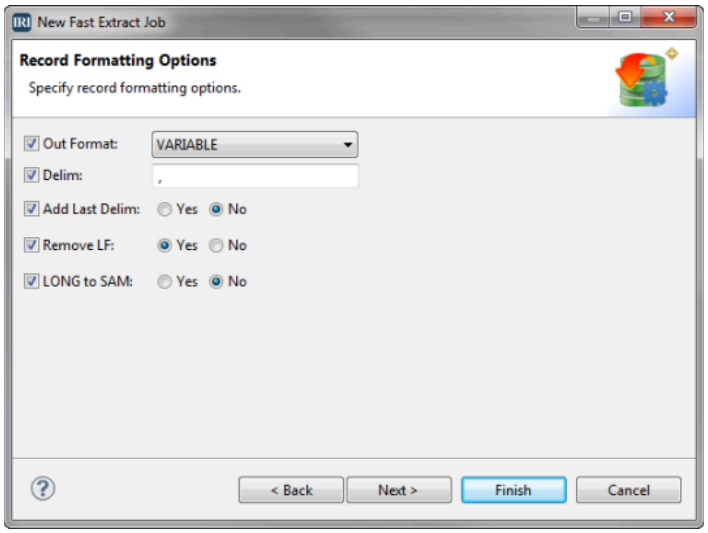
After identifying the job script and adding comments on the initial page, the wizard has you define the details about your database connection, which can also be abstracted through the "FACT Manager" utility for added security.

The SQL query is specified on this page, along with several split function parameters for parallel extraction.

File Naming and Loading Options

Here you specify the default filenames for extracted data, and the directories where data and metadata files get written. You can also determine unload subset sizes, and downstream load methodology.





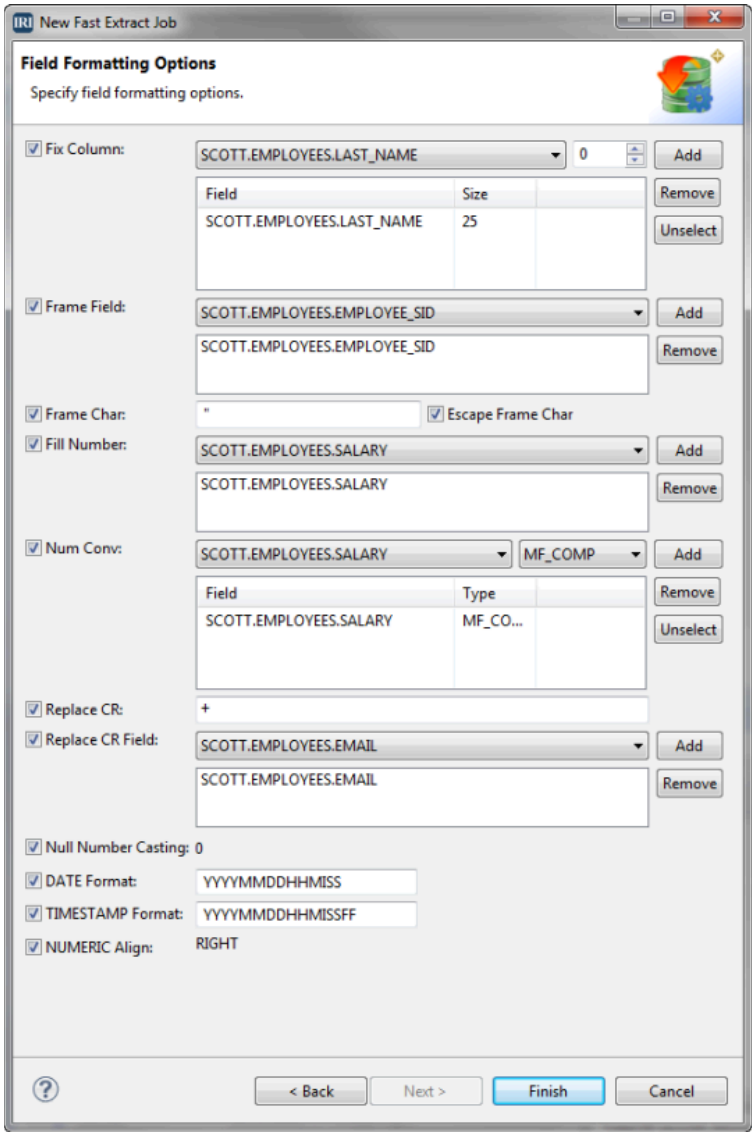
Record Formatting Options

Although FACT analyzes the columns as it performs the extraction and sets the data type and column lengths automatically, you can still modify the field and record formats in the extracted file.

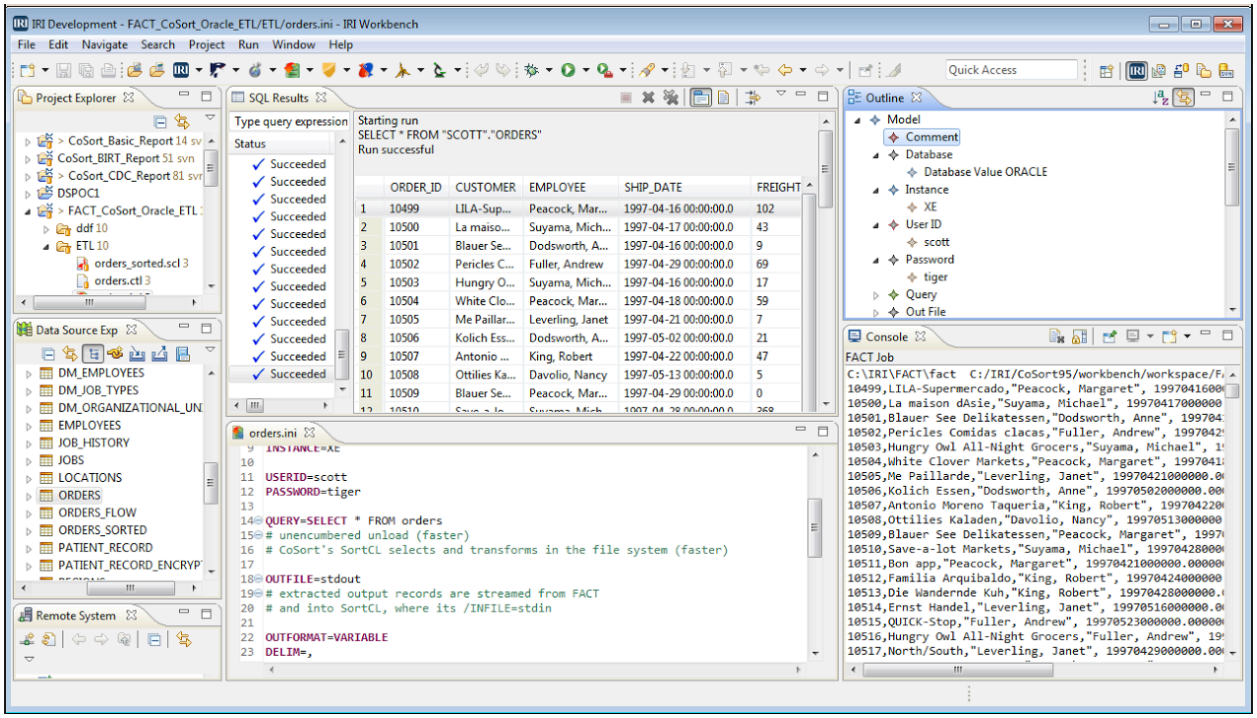
Field Formatting Options

This page continues with more formatting options for the extracted data. For even more granular control of the data after extraction, use CoSort SortCL program parameters for:

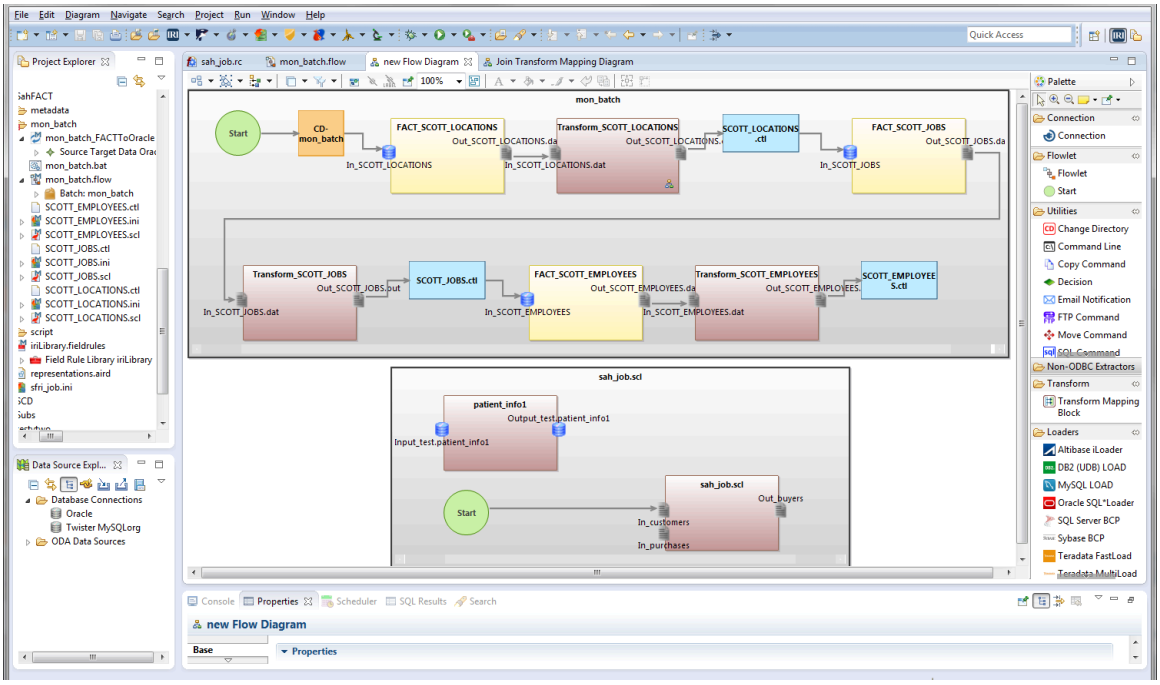
- field repositioning and resizing
- data type conversion
- endian conversion
- pre-load sorting
- static or running aggregation
- cross-calculation
- lookup transformations
- row-column pivoting
- data (PII) masking
- date masking
- inserting labels or strings
- assigning index keys
- data cleansing
- find and replace
- report formatting
- file format conversion



At the end of the job, the FACT configuration file is ready to go, and fully modifiable in any text editor, or within the syntax-aware FACT editor in the IRI Workbench. Here you can see the source and target data with the FACT configuration file in the Workbench:

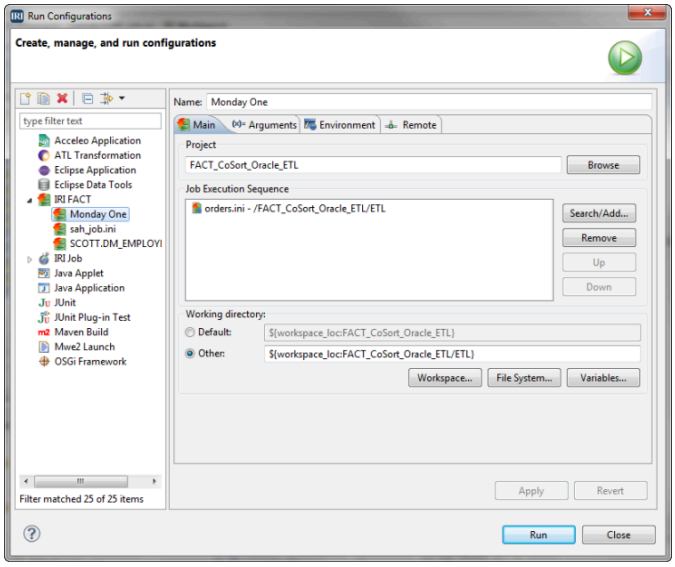


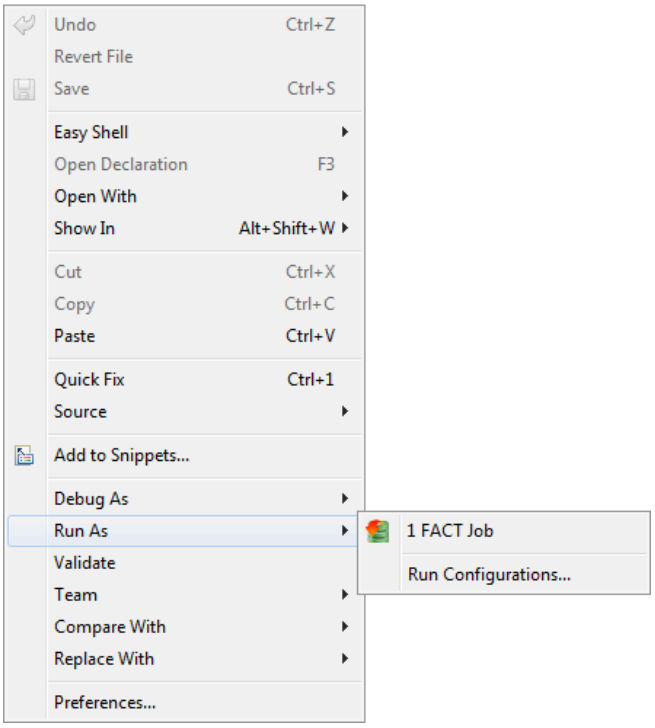
The IRI Workbench is also where Voracity data warehouse architects can use FACT's optional downstream Transform and Load metadata in CoSort-powered data warehouse ETL, DB migration, and VLDB reorg operations, and design those jobs to run directly with FACT:



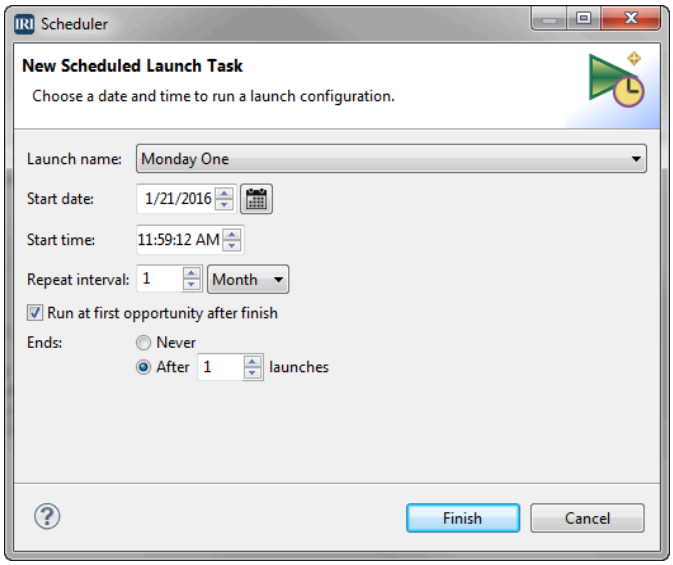
FACT Execution

Launch or schedule the extraction job from the 'Run As' menu in IRI Workbench (top navigation or right-click), or from an ETL or offline or reorg job in IRI Voracity, etc.





FACT jobs can run in the IRI Workbench from the project explorer, a Voracity ETL workflow, the offline reorg wizard, or as an automated task with rules defined in the onboard task scheduler:



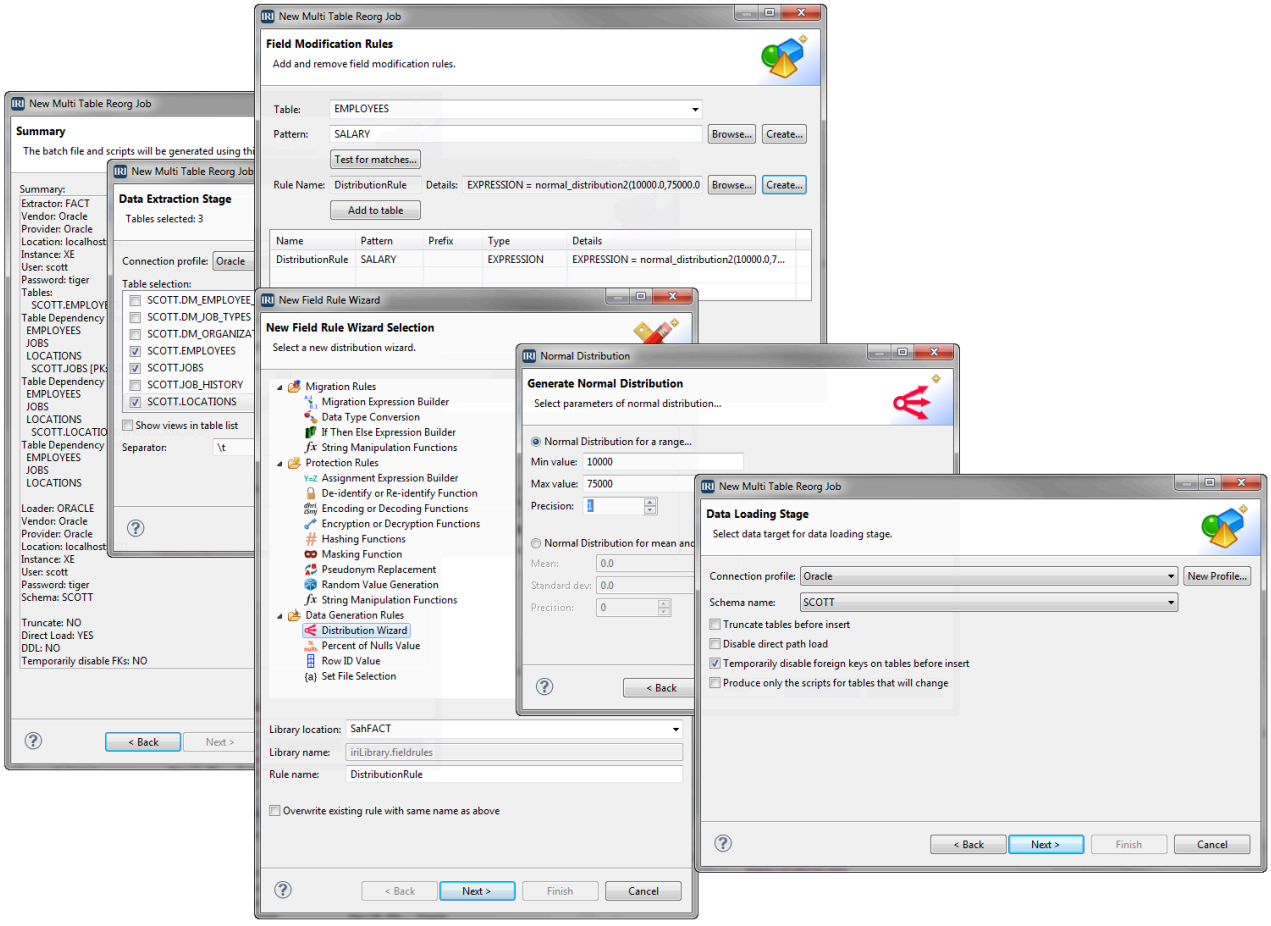
As a command-line utility routinely called into shell (batch) scripts, FACT jobs can also be invoked from any third-party scheduler, including the Oracle DBMS (Jobs) Scheduler and the Universal Automation Center from Stonebranch.

Using FACT in Database Reorgs

As database table and index sizes increase, data becomes more fragmented and query response slows. To improve database operating efficiency, regular table reorganization is required. The *Multi-Table Reorg Job ...* wizard in the IRI Workbench builds a step-by-step, high-performance “Unload-Sort-Reload” process for one or more tables at once, using:|

- IRI FACT for bulk table unloads
- IRI CoSort for their reordering
- the target database’s load utility for pre-sorted, bulk loads

ODBC select and insert options are also available for smaller-scale or finer-tuned operations.



At the end of the wizard, the job scripts needed to reorg the selected tables are created. The jobs can run anywhere the chosen tools are licensed, and be invoked from the GUI, command line, or batch script (which the wizard also produces). Database users are unaffected in the offline reorg method, although reloads or ODBC updates can alter tables in use.

The versatile *New Field Rule Wizard Selection* page guides you through the selection and creation of a specific rule you want to apply to your data. The *Summary* provides the opportunity to review your job setup before actually generating the batch file and scripts.

FACT Specifications

- Uses native DB connection protocols
- Multiple parallel unload methods
- Flat-file output
- Automatic DDF and load config file creation
- ETL documentation and demo project
- Column, date, and COBOL type conversion
- Special COB, LOB, and LONG extract options
- Simple config and SQL SELECT syntax
- New job wizard in Eclipse GUI
- Offline reorg wizard in Eclipse

FACT Licensing

FACT is licensed for perpetual use between $5K and $50K as a standalone product, based on the number of threads you need to leverage on any given hostname. Subscription licenses are available when used with the IRI Voracity data management (ETL) platform. Discounts apply on multiple copies licensed at once and to 501(c)3 non-profit institutions. All licenses include documentation and at least one year of technical support.

INNOVATIVE ROUTINES INTERNATIONAL (IRI), INC.

2194 Highway A1A
Melbourne, FL 32937 USA
Phone +1 321-777-8889
iri.com/fact