

ETL Tool Questionnaires

Author: Dhananjay Patil
Organization: Evaltech, Inc.
Evaltech Research Group,
Data Warehousing Practice.
Date: 08/13/04
Email: erg@evaltech.com

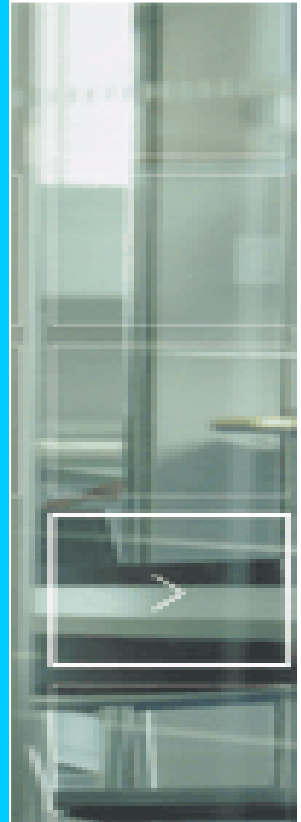


Abstract:

This white paper includes ETL tool related some questions and answers, which clears the general concepts of ETL tool. As example, list of databases and ERP systems supports to ETL tool. How ETL tool play a role while loading? What is the Latest version of ETL tool? What are the built in transformations? This article will be help full for describing the importance of ETL tool, and also currently in the markets how it is expanding.

Intellectual Property / Copyright Material

All text and graphics found in this article are the property of the Evaltech, Inc. and cannot be used or duplicated without the express written permission of the corporation through the Office of Evaltech, Inc.



Questions and Concerns

Q1	List the Databases, the Tool can support?
	Oracle, DB2, Teradata, Informix, Microsoft SQL Server
Q2	List the ERP Systems, the Tool can support?
	The biggest ERP vendors include Baan, J.D. Edwards, Oracle, PeopleSoft, and SAP.
Q3	What if we have source that is not supported by your tool?
	Create a Flat file and upload it .
Q4	How your Tool can connect to database under Unix?
	By using an ODBC connection as example ODBC—AIX,HP-UX aCC Enabled (PA-RISC and IPF),Sun solaris,Linux
Q5	List the built-in transformation objects?
	<ul style="list-style-type: none"> • SQL steps (SELECT and INSERT) • DB2 utilities (LOAD, RUNSTATS, EXPORT, REORG) • Data cleansing transformers (Java, stored procedures, user-defined functions) • Statistical transformers (Java and stored procedures) • Replication APPLY • OLAP steps (LOAD, CALC, and more) • File programs (FTP, submit OS/390 JCL, and more) • Interoperability steps • ETI, Data Stage, OIS, Integrity, Trillium
Q6	Does the Tool support de-normalization? If yes How?
	Yes. Transfer data from different source table to one target table. ETL tool will create another key field i.e. called as a indexed field if the same key field exist in no of table.
Q7	Does the Tool support and Star Schema? If yes How?
	Yes, in following manner,

	<p>Star Schema Builder Within the Data Warehouse Center, the warehouse schema modeler is a specialized tool for generating and storing schema associated with a data warehouse. Any schema resulting from this process can be passed easily as metadata to an OLAP tool.</p>
Q8	<p>Does it have Scheduler?</p>
	<p>Yes which available in Work Flow Manager</p>
Q9	<p>Can Tool Sort Data?</p>
	<p>Yes by using sorter transformation</p>
Q10	<p>How the Tool manages the Meta Data?</p>
	<p>The Information Catalog Center helps end users find, understand, and access available information stored in the data warehouse.</p> <p>Manage business metadata for end users</p> <ul style="list-style-type: none"> • Web or Windows end user interface • Describes information in business terms • Catalog search engine • Documents currency • Launches tool which renders information <p>Supports information registration and sharing</p> <ul style="list-style-type: none"> • Supports any type of information object • Databases, cubes, queries, reports, charts, spreadsheets, and web pages • Maintains object metadata: name, description, owner, currency, associated tools, and more <p>Supports grouping by categories and varying user authorities</p> <ul style="list-style-type: none"> • Supports automatic registration of information objects • Automatic population and synchronization with Data Warehouse Center metadata • Pre-built metadata interchange • QMF, Brio, Business Objects, Cognos, Hyperion, and desktop tools <p>The real value in the information catalog is the data it contains. The information catalog supports automatic registration of information objects. You can populate the information catalog from the Data Warehouse Center. The administrator can select the objects for publishing and automatic synchronization. The information catalog can be automatically populated by many tools including QMF, Brio, Business Objects, Cognos, Hyperion Essbase, and popular desktop tools such as Microsoft Excel, Lotus 1-2-3, and others.</p>

Q11	Is the Meta data stored in RDBMS? If not then how?
	No, Meta data stored in repository.
Q12	Is it Engine Driven or Code Generated? Explain its advantages?
	<p>Yes, The developer skill level required to use these tools successfully is precisely that same set of skills that is in short supply to understand the source environments in the first place. So, rather than relieving the skill and manpower burden, the tools often just aggravate this problem. And Since the operational source systems typically run on different schedules and transaction volume rates. These execution and availability schedule variances can make it extremely difficult to extract, integrate, merge, or compare and reconcile multiple source system records prior to transformation. Logically defining a single batch process to handle these timing differences is extremely complex. Data can only be staged to temporary files or tables by creating multiple, dependent extraction/transformation processes and then creating subsequent transformation processes using the interim results. The design burden for staging data in this fashion falls on the developer with no assistance from the tool to help determine how to partition the process. Furthermore, the developer also has to define the operational procedures, such as a job stream with multiple procedures to coordinate the various extraction/transformation steps. This extra work, outside the context of the tool, undermines the productivity benefits to be derived.</p> <p>Finally, none of the code-generating tools was designed for the now de facto standard Windows user interface.</p>
Q13	Does the tool support bulk loading? If yes how?
	Yes, data transfer from source to target at a time instead of one by one records. But bulk loading is depends on database system. As example SQL server supports bulk loading while Oracle server does not support.
Q14	How does the tool supports scalability?
	<p>As far as scalability is concerned, organizations should concern with two characteristics: massive data throughput and fast processing. Parallel processing capabilities in every component of the tool, as well as clustering and load balancing functions for deployments of multiple server instances.</p> <p>Similarly, an enterprise-class ETL tool must provide a development environment that can accommodate multiple developers—or multiple teams of developers—at the same time. Enterprise scale use assumes that multiple developers will share the design environment, creating and updating ETL objects and projects simultaneously. Look for check-in/check-out and versioning for both objects and projects. To enable collaboration and project management, there should be central repository containing ETL objects and projects (not just metadata). Geographically dispersed teams need Internet-based access to the repository.</p>

Q15	Does the tool support incremental load? If yes how?
	<p>Yes. Basically there are two ways. 1) Extract All Records: Extract source data from all operational records, regardless of whether any data values have changed since the last ETL load or not</p> <p>2) Extract Deltas Only: Extract source data only from those operational records in which some data values have changed since the last ETL load (“net change”)</p>
Q16	How does the tool support the security model?
	Only authorized person can start services provided in the tool.
Q17	Does the tool have debugging capability? If yes how?
	Yes. We can do the debugging at the time of mappings and scheduling the job.
Q18	Does the tool have Backup and recovery features?
	Yes
Q19	Does the tool control the version? If yes how?
	Yes (informatica version 7.1 supports). When every time create a mapping or workflow, there is a need to do checking to affect the changes on other machine. This checking means creating or controlling the version.
Q20	Does jobs/sessions get locked frequently while loading?
	Yes
Q21	How the tool can handle Rejects?
	Updatestratagy and Filter transformation even there is option also in source qualifiers
Q22	List the Data Modeling tools it supports?
	<p>AllFusion ERwin Data Modeler Repository Erwin Visio Advantage Gen COOL:Gen IEF Composer COOL:Enterprise Key:Enterprise Application Development Workbench (ADW) COOL:BusinessTeam GroundWorks Analyst COOL:DBA Terrain for DB2</p>

	DBA COOL:Xtras Mapper TerrainMap ER/Studio XDE (via MDL) Rose Data Modeler AppMaster Designer AppMaster Designer Excelerator II Designer Designer 2000 System Architect ProActivity Select SE Systems Engineer Silverrun RDM PowerDesigner / CDM PowerDesigner / PDM PowerDesigner / CDM DataArchitect / CDM S-Designer / CDM PowerDesigner / PDM DataArchitect / PDM S-Designer / PDM IE:Advantage
Q23	Does it have integration with other products? If yes list the products?
	Yes.
Q24	Does the tool support XML, Web logs, MQSeries?
	Yes for XML files XML source qualifiers and for MQSeries MQSource qualifiers.
Q25	What Operating System does it support? List all of them?
	Windows NT, Unix, IBM Series
Q 26	What is the Tool's backup and recovery strategy?
	Since tool can backup a repository.
Q27	How long has the product been on the market?
	1990s
Q28	Do health care companies use your product? If yes please list the names and their case study? Can we get the reference?
	?
Q29	How you take care of product upgrades?
	?

Q30	Do you have existing customer's feedback?										
	?										
Q31	What is your Licensing arrangement? Is it per user/CPU?										
	?										
Q32	List the infrastructure requirements? Software and Hardware?										
	?										
Q33	What is the latest version? What is the release date?										
	<table> <tr> <td>Business Objects (Data Integrator)</td> <td>6.0 for CWM 1.0</td> </tr> <tr> <td>ActaWorks</td> <td>5.x for CWM 1.0</td> </tr> <tr> <td>DB2 Warehouse Manager</td> <td>7.2</td> </tr> <tr> <td>ETL Studio for CWM</td> <td>1.0</td> </tr> <tr> <td>PowerCenter</td> <td>5.1 to 7.x</td> </tr> </table>	Business Objects (Data Integrator)	6.0 for CWM 1.0	ActaWorks	5.x for CWM 1.0	DB2 Warehouse Manager	7.2	ETL Studio for CWM	1.0	PowerCenter	5.1 to 7.x
Business Objects (Data Integrator)	6.0 for CWM 1.0										
ActaWorks	5.x for CWM 1.0										
DB2 Warehouse Manager	7.2										
ETL Studio for CWM	1.0										
PowerCenter	5.1 to 7.x										