

ETL Basics

Lesson 4: ETL Tools

Lesson Objectives

- On completion of this lesson on ETL basics, you will be able to understand:
 - Consideration of ETL tool
 - Different ETL tools



The ETL Process

- Extract
 - Extract relevant data
- Transform
 - Transform data to DW format
 - Build keys, etc.
 - Cleansing of data
- Load
 - Load data into DW
 - Build aggregates, etc

Criteria for tool identification

- Criteria for Identifying Tools
 - The Source System Platform and Database.
 - - Tools cannot access all types of data source on all types of Computing platforms
 - Data Size
 - - Tools need to handle desired volume and type of data. E.g. Structured and unstructured data.
 - Functionality required
 - - Tools have build in functionalities
 - Cost of the tool



Copyright © Capgemini 2016. All Rights Reserved. 4

Various tools are available for extraction process. Selecting a suitable tools becomes an important criteria for the success of the warehouse implementation.

The tools should be capable of selecting the data from various sources without user interface. Connectors are available for extracting data from various sources from various computing platforms.

Some of the tools have built-in extraction which reduces the manual coding activity.

SAMPLE Extraction Tools

- Extraction Tools include
- Apertus Carleton. Passport
- Evolutionary Technologies. ETL Extract.
- Platinum. InfoPump

These are the some of the Industry standard tools which are used for extracting data from either single or multiple sources.

Carleton, Passport

Users enter extraction and transformation parameters, and Data is filtered against domains and ranges of legal values.

Evolutionary Technology (ETI), EXTRACT

Users write transformation rules. Data is filtered against domains and ranges of legal values and compared to other data structures

Platinum, InfoPump

A data pump product designed to extract data from several mainframe and client server platforms, perform some filtering and transformation, and distribute and load to another mainframe platform database.

Requires InfoHub for most services.

The extraction uses custom code modules. This is a client/server based tool.

In addition there some more tools available in the market such as

- Information Discovery ie., IDI

 - Low-end rule discovery

Oracle, Symmetric Replicator

A data replication product designed to extract data from several platforms, perform some filtering and transformation, and distributes and loads data in to one or more database or databases.

In addition there are some more tools available in the market such as

- Information Discovery, IDI

 - Low-end rule discovery

Oracle, Symmetric Replicator

A data replication product designed to extract data from several platforms, perform some filtering and transformation, and distributes and loads data in to one or more database or databases.

Sample ETL Tools

- Teradata Warehouse Builder from Teradata
- DataStage from IBM
- SAS System from SAS Institute
- Power Mart / Power Center from Informatica
- Sagent Solution from Sagent Software
- Hummingbird Genio Suite from Hummingbird Communication
- Abinitio
- Oracle Warehouse Builder
- Talend
- Pentaho



Copyright © Capgemini 2010. All Rights Reserved. 7

You might go to the vendors' web sites to find a good demo to show your students.

Sample Scheduling Tools

- AutoSys
- Control-M
- Flux
- IBM Workload Scheduler (TWS)

You might go to the vendors' web sites to find a good demo to show your students.

Sample Reporting Tools

- SAP- Business Object
- IBM- Cognos
- Microsoft SQL Server Reporting Services -SSRS
- Microstrategy
- Oracle Business Intelligence Enterprise Edition

You might go to the vendors' web sites to find a good demo to show your students.

Summary

- In this module, you learned about the following:
 - Consideration for selecting ETL tool
 - Different ETL tools



Add the notes here.