

ETL :-

- extract -- source info(flat file ,data base tables)
- transform d source dat which is in different format into a common format
- load - in to target(flat file ,data base tables)

Informatica- a data integration compny

etl product - power center

server:

repository service nd integration service

Cleint:

-Designer, workflow,manager,repository manager,monitor

SQL-

orcale,microsoft sql server,teradata etc.

.types of joins:-

.order by , .groupby, .where, .having, .select, .insert, .update, .delete, .truncate

Sql:

source minus target - zero

target minus source - zero

Source;:- we need to write the sql on the transformation that is required.

Target: the value which is populaytd by ETL job

Data Warehousing Concepts and terms:

2 terms that are of great imporatnce in undrstanding of data warehousing concepts are

Operational Data:.....

It is the data is used to run a business. this data is what is typically

stored.retrived and updated by online transaction processing(OLTP) system.

Operatinal data is stored in a relational database.but may be stored inlegacy.hierrchical or flat file formts as well

Informational data:....

It is stored on format that makes analysis much easier.analysis can be in the form of decission support,report gneration.executive info systems & more in depth statical analyse

DEfintin::: A strcture extensible envt designed for the analysis of non-volatile

data.logically & physically transformed from multiple source application to align with business structr.

updated & maintained for a long time period

Q)Why data warehouse in to picture?

- Data warehouse implementation solved user's below prob:

- 1- I cant find the data I need
- 2- I cant get the data i need
- 3- I cant Use the data I need

Q)Differnce bewn OLTP & Datawarehouse??

-Perfomance sensitive	Relaxed
-Few Recoreds accessed at atime(tens) read/update access	Large volumes access at a time(millions) Mostly read Access
data redundancy is absent	Redudancy is present
Database size(100 mb - 100 Gb)	Database size(100 gb - few terabyte)
Application oriented	Subject oriented
Detailed data	summarized & Refined data
Isolated data	Integrated data
NOrmal user	knowledge users
Current up to date	Snap shot data

ETL overview::

- Extract Transform and Load is used to populate a data warehouse
- Extract is where data is pulled from source systems
 - Sql Connect over networks
 - Flat files
 - Transacation messaging((MSMQ))
- Transformation can be the most complex part of data warehousing
 - convert to numbers
 - Apply business logic in this stage
- Load is where data is loaded into the data warehouse
 - Sequential or bulk loading

Total Course is :-

- ✓ Informatica
 - ✓ DWH
 - ✓ Data Modelling
 - ✓ SQL
 - ✓ Basics of PLsql
 - ✓ UNIX Command
 - ✓ What are the BI tools are we are using?
 - ✓ Manual Testing, SDLC,STLC, BUG DEFECT LIFE CYCLE..
-
- ❖ Theory class
 - ❖ Practical sessions
 - ❖ Interview point of view Q/A
 - ❖ Real time Project
 - ❖ Project Related Q/A