

# **Extract, transform, load (ETL) Interview Questions And Answers Guide.**



**Global Guideline.**

<https://globalguideline.com/>



# Extract, transform, load (ETL) Job Interview Preparation Guide.

## Question # 1

Explain a mapping, session, worklet, workflow, maplet?

### Answer:-

A mapping represents dataflow from sources to targets.

A maplet creates or configures a set of transformations.

A workflow is a set of instruction that tell the Informatica server how to execute the tasks.

A worklet is an object that represents a set of tasks.

A session is a set of instructions that describe how and when to move data from sources to targets.

[Read More Answers.](#)

## Question # 2

Explain Where do we use semi and non additive facts?

### Answer:-

Additive: A measure can participate arithmetic calculations using all or any dimensions.

Ex: Sales profit

Semi additive: A measure can participate arithmetic calculations using some dimensions.

Ex: Sales amount

Non Additive: A measure can't participate arithmetic calculations using dimensions.

Ex: temperature

[Read More Answers.](#)

## Question # 3

Suppose we have some 10,000 odd records in source system and when load them into target. How do we ensure that all 10,000 records that are loaded to target doesn't contain any garbage values?

### Answer:-

You can verify the session log and ensure the source fetched and target loaded details

[Read More Answers.](#)

## Question # 4

How to calculate fact table granularity?

### Answer:-

Granularity, is the level of detail in which the fact table is describing, for example if we are making time analysis so the granularity maybe day based - month based or year based

[Read More Answers.](#)

## Question # 5

Explain What is ETL process? How many steps ETL contains explain with example?

### Answer:-

ETL is extraction, transforming, loading process, you will extract data from the source and apply the business rule on it then you will load it in the target the steps are:

1-define the source(create the odbc and the connection to the source DB)

2-define the target (create the odbc and the connection to the target DB)

3-create the mapping ( you will apply the business rule here by adding transformations, and define how the data flow will go from the source to the target )

4-create the session (its a set of instruction that run the mapping, )

5-create the work flow (instruction that run the session)

[Read More Answers.](#)



### Question # 6

Tell me can we override a native sql query within Informatica? Where do we do it? How do we do it?

#### Answer:-

Yes, we can override a native sql query in source qualifier and lookup transformation.

In lookup transformation we can find "Sql override" in lookup properties. by using this option we can do this.

[Read More Answers.](#)

### Question # 7

Explain What are the various methods of getting incremental records or delta records from the source systems?

#### Answer:-

One foolproof method is to maintain a field called 'Last Extraction Date' and then impose a condition in the code saying 'current\_extraction\_date > last\_extraction\_date'.

First Method: If there is a column in the source which identifies the record inserted date. Then it will be easy to put a filter condition in the source qualifier.

Second Method: If there is no record in the source to identify the record inserted date. Then we need to do a target lookup based on the primary key and determine the new record and then insert.

[Read More Answers.](#)

### Question # 8

Tell me what is Full load & Incremental or Refresh load?

#### Answer:-

Full Load: completely erasing the contents of one or more tables and reloading with fresh data.

Incremental Load: applying ongoing changes to one or more tables based on a predefined schedule

Refresh Load: the table will be truncated and data will be loaded again. Here we use to load static dimension table or type tables using this method.

Incremental Load: It is a method to capture on the newly created or updated record. Based upon the flag or Date this load will be performed.

Full Load: when we are loading the data for first time, either it may be a base load or history all the set of records will be loaded at a stretch depends upon the volume.

[Read More Answers.](#)

### Question # 9

Explain What is the difference between Power Center & Power Mart?

#### Answer:-

PowerCenter - ability to organize repositories into a data mart domain and share metadata across repositories.

PowerMart - only local repository can be created.

[Read More Answers.](#)

### Question # 10

Explain the various test procedures used to check whether the data is loaded in the backend, performance of the mapping, and quality of the data loaded in INFORMatica?

#### Answer:-

the best procedure to take a help of debugger where we monitor each and every process of mappings and how data is loading based on conditions breaks.

[Read More Answers.](#)

### Question # 11

Explain the difference between keylookup and joiner?

#### Answer:-

Explain your self

[Read More Answers.](#)

### Question # 12

Explain and Compare ETL & Manual development?

#### Answer:-

ETL - The process of extracting data from multiple sources.(ex. flat files,XML, COBOL, SAP etc) is more simpler with the help of tools.

Manual - Loading the data other than flat files and oracle table need more effort.

ETL - High and clear visibility of logic.

Manual - complex and not so user friendly visibility of logic.

ETL - Contains Meta data and changes can be done easily.

Manual - No Meta data concept and changes needs more effort.

ETL - Error handling, log summary and load progress makes life easier for developer and maintainer.

Manual - need maximum effort from maintenance point of view.

ETL - Can handle Historic data very well.

Manual - as data grows the processing time degrades.

These are some differences b/w manual and ETL development.

[Read More Answers.](#)

### Question # 13

Explain What are the different versions of Informatica?

**Answer:-**

Here are some popular versions of Informatica.

Informatica Powercenter 4.1, Informatica Powercenter 5.1, Powercenter Informatica 6.1.2, Informatica Powercenter 7.1.2, Informatica Powercenter 8.1, Informatica Powercenter 8.5, Informatica Powercenter 8.6.

[Read More Answers.](#)

**Question # 14**

What is latest version of Power Center / Power Mart?

**Answer:-**

The Latest Version is 7.2  
informatica 8.0

[Read More Answers.](#)

**Question # 15**

Explain What is ODS (operation data source)?

**Answer:-**

ODS - Operational Data Store.

ODS Comes between staging area & Data Warehouse. The data in ODS will be at the low level of granularity.

Once data was populated in ODS aggregated data will be loaded into EDW through ODS.

ODS: ODS is also a similar small DWH which will help analyst to analyze the business. It will have data for less number of days. Generally it will be around 30-45 days. Like DWH here also we will primary keys will be generated, error and reject handling will be done.

[Read More Answers.](#)

**Question # 16**

How to use procedural logic inside Informatica? If yes how, if not how can we use external procedural logic in Informatica?

**Answer:-**

We can use advanced external transformation. For more detail you can refer the manual of Informatica transformation guide in that advanced external transformation. You can use C++ language on Unix and C++, VB, VC++ on Windows server.

[Read More Answers.](#)

**Question # 17**

Explain Is there any way to read the MS Excel Data directly into Informatica? Like IS there any possibilities to take excel file as target?

**Answer:-**

We can't directly import the XML file in Informatica.

We have to define the Microsoft Excel ODBC driver on our system. and define the name in excel sheet by defining ranges.

Then in Informatica open the folder using sources -> import from database -> select excel ODBC driver -> connect -> select the excel sheet name.

[Read More Answers.](#)

**Question # 18**

Explain What are snapshots? What are materialized views & where do we use them? What is a materialized view log?

**Answer:-**

Snapshots are read-only copies of a master table located on a remote node which is periodically refreshed to reflect changes made to the master table. Snapshots are mirror or replicas of tables.

Views are built using the columns from one or more tables. The Single Table View can be updated but the view with multi table cannot be updated.

A View can be updated/deleted/inserted if it has only one base table if the view is based on columns from one or more tables then insert, update and delete is not possible.

Materialized view

A pre-computed table comprising aggregated or joined data from fact and possibly dimension tables. Also known as a summary or aggregate table.

[Read More Answers.](#)

**Question # 19**

Explain Techniques of Error Handling - Ignore, Rejecting bad records to a flat file, loading the records and reviewing them (default values)?

**Answer:-**

Rejection of records either at the database due to constraint key violation or the Informatica server when writing data into target table. These rejected records we can find in the badfiles folder where a reject file will be created for a session. We can check why a record has been rejected. And this bad file contains first column a row indicator and second column a column indicator.

These row indicators are of four types

D- valid data,

O- overflowed data,

N- null data,

T- Truncated data,

And depending on these indicators we can change to load data successfully to target.

[Read More Answers.](#)

**Question # 20**

Explain What are the various transformations available?



### **Answer:-**

Aggregator Transformation  
Expression Transformation  
Filter Transformation  
Joiner Transformation  
Lookup Transformation  
Normalizer Transformation  
Rank Transformation  
Router Transformation  
Sequence Generator Transformation  
Stored Procedure Transformation  
Sorter Transformation  
Update Strategy Transformation  
XML Source Qualifier Transformation  
Advanced External Procedure Transformation  
External Transformation

[Read More Answers.](#)

### **Question # 21**

How to extract SAP data Using Informatica? What is ABAP? What are IDOCS?

### **Answer:-**

SAP Data can be loaded into Informatica in the form of Flat files.  
Condition :  
Informatica source qualifier column sequence must match the SAP source file.

[Read More Answers.](#)

### **Question # 22**

Explain What is partitioning? What are the types of partitioning?

### **Answer:-**

If you use PowerCenter, you can increase the number of partitions in a pipeline to improve session performance. Increasing the number of partitions allows the Informatica Server to create multiple connections to sources and process partitions of source data concurrently.  
When you create a session, the Workflow Manager validates each pipeline in the mapping for partitioning. You can specify multiple partitions in a pipeline if the Informatica Server can maintain data consistency when it processes the partitioned data.  
When you configure the partitioning information for a pipeline, you must specify a partition type at each partition point in the pipeline.  
The partition type determines how the Informatica Server redistributes data across partition points.  
The Workflow Manager allows you to specify the following partition types:  
Round-robin partitioning. The Informatica Server distributes data evenly among all partitions. Use round-robin partitioning where you want each partition to process approximately the same number of rows.  
For more information, see Round-Robin Partitioning.  
Hash partitioning. The Informatica Server applies a hash function to a partition key to group data among partitions. If you select hash auto-keys, the Informatica Server uses all grouped or sorted ports as the partition key. If you select hash user keys, you specify a number of ports to form the partition key. Use hash partitioning where you want to ensure that the Informatica Server processes groups of rows with the same partition key in the same partition. For more information, see Hash Partitioning.  
Key range partitioning. You specify one or more ports to form a compound partition key. The Informatica Server passes data to each partition depending on the ranges you specify for each port. Use key range partitioning where the sources or targets in the pipeline are partitioned by key range. For more information, see Key Range Partitioning.  
Pass-through partitioning. The Informatica Server passes all rows at one partition point to the next partition point without redistributing them. Choose pass-through partitioning where you want to create an additional pipeline stage to improve performance, but do not want to change the distribution of data across partitions.

[Read More Answers.](#)

### **Question # 23**

Explain Do we need an ETL tool? When do we go for the tools in the market?

### **Answer:-**

ETL Tool:  
It is used to Extract(E) data from multiple source systems(like RDBMS,Flat files,Mainframes,SAP,XML etc) transform(T) them based on Business requirements and Load(L) in target locations.(like tables,files etc).  
Need of ETL Tool:  
An ETL tool is typically required when data scattered across different systems.(like RDBMS,Flat files,Mainframes,SAP,XML etc).

[Read More Answers.](#)

### **Question # 24**

Explain What is Informatica Metadata and where is it stored?

### **Answer:-**

Informatica Metadata is data about data which stores in Informatica repositories.

[Read More Answers.](#)

### **Question # 25**

Explain What is a staging area? Do we need it? What is the purpose of a staging area?

### **Answer:-**

Data staging is actually a collection of processes used to prepare source system data for loading a data warehouse. Staging includes the following steps:



Source data extraction, Data transformation (restructuring),  
Data transformation (data cleansing, value transformations),  
Surrogate key assignments

[Read More Answers.](#)

### **Question # 26**

Explain What is a three tier data warehouse?

#### **Answer:-**

A data warehouse can be thought of as a three-tier system in which a middle system provides usable data in a secure way to end users. On either side of this middle system are the end users and the back-end data stores.

[Read More Answers.](#)

### **Question # 27**

Explain When do we Analyze the tables? How do we do it?

#### **Answer:-**

The ANALYZE statement allows you to validate and compute statistics for an index, table, or cluster. These statistics are used by the cost-based optimizer when it calculates the most efficient plan for retrieval. In addition to its role in statement optimization, ANALYZE also helps in validating object structures and in managing space in your system. You can choose the following operations: COMPUTE, ESTIMATE, and DELETE. Early version of Oracle7 produced unpredictable results when the ESTIMATE operation was used. It is best to compute your statistics.

EX:

```
select OWNER,
sum(decode(nvl(NUM_ROWS,9999), 9999,0,1)) analyzed,
sum(decode(nvl(NUM_ROWS,9999), 9999,1,0)) not_analyzed,
count(TABLE_NAME) total
from dba_tables
where OWNER not in ('SYS', 'SYSTEM')
group by OWNER
```

[Read More Answers.](#)

### **Question # 28**

How to call shell scripts from informatica?

#### **Answer:-**

Specify the Full path of the Shell script the "Post session properties of session/workflow".

[Read More Answers.](#)

### **Question # 29**

Explain What is the metadata extension?

#### **Answer:-**

Informatica allows end users and partners to extend the metadata stored in the repository by associating information with individual objects in the repository. For example, when you create a mapping, you can store your contact information with the mapping. You associate information with repository metadata using metadata extensions.

Informatica Client applications can contain the following types of metadata extensions:

Vendor-defined. Third-party application vendors create vendor-defined metadata extensions. You can view and change the values of vendor-defined metadata extensions, but you cannot create, delete, or redefine them.

User-defined. You create user-defined metadata extensions using PowerCenter/PowerMart. You can create, edit, delete, and view user-defined metadata extensions. You can also change the values of user-defined extensions.

[Read More Answers.](#)

### **Question # 30**

Explain What is Full load & Incremental or Refresh load?

#### **Answer:-**

Full Load: completely erasing the contents of one or more tables and reloading with fresh data.

Incremental Load: applying ongoing changes to one or more tables based on a predefined schedule.

first time what we are loading the data is called initial load or full load.

secondtime or modified data what we are loading is called as incremental load or delta load.

[Read More Answers.](#)

### **Question # 31**

Explain the different Lookup methods used in Informatica?

#### **Answer:-**

1. Connected lookup

2. Unconnected lookup

Connected lookup will receive input from the pipeline and sends output to the pipeline and can return any number of values.it does not contain return port.

Unconnected lookup can return only one column. it containn return port.

[Read More Answers.](#)



### **Question # 32**

Explain the process of extracting data from source systems, storing in ODS and how data modeling is done?

#### **Answer:-**

There are various ways of Extracting Data from Source Systems. For example, you can use a DATA step, an Import Process. It depends with your input data styles. What kind of File/database it is residing in. Storing ur data in an ODS can be done thru an ODS stmt/export stmt/FILE stmt, again which depends on the file & data format. You want your output to be in.

[Read More Answers.](#)

### **Question # 33**

Explain if a flat file contains 1000 records how can i get first and last records only?

#### **Answer:-**

By using Aggregator transformation with first and last functions we can get first and last record.

[Read More Answers.](#)

### **Question # 34**

Explain the difference between etl tool and olap tools?

#### **Answer:-**

ETL tool is ment for extraction data from the legecy systems and load into specified data base with some process of cleansing data.

ex: Informatica, data stage ....etc

OLAP is ment for Reporting purpose. in OLAP data available in Multidimensional model. so that u can write smple query to extract data fro the data base.

ex: Businee objects, Cognos....etc

[Read More Answers.](#)

### **Question # 35**

Explain What are the modules in Power Mart?

#### **Answer:-**

1. PowerMart Designer
2. Server
3. Server Manager
4. Repository
5. Repository Manager

[Read More Answers.](#)

### **Question # 36**

What are the various tools? Explain Name a few?

#### **Answer:-**

A few more ...

- Cognos Decision Stream
- Oracle Warehouse Builder
- Business Objects XI (Extreme Insight)
- SAP Business Warehouse
- SAS Enterprise ETL Server

Tools of SAS are :-

- SAS Information delivery portal.
- SAS data integration studio.
- Business Intelligence.
- SAS Enterprise Guide.

[Read More Answers.](#)

### **Question # 37**

How can we determine what records to extract?

#### **Answer:-**

When addressing a table some dimension key must reflect the need for a record to get extracted. Mostly it will be from time dimension (e.g. date >= 1st of current mth) or a transaction flag (e.g. Order Invoiced Stat). Foolproof would be adding an archive flag to record which gets reset when record changes.

[Read More Answers.](#)

## Data Warehouse Most Popular Interview Topics.

- 1 : [Data Warehousing Frequently Asked Interview Questions and Answers Guide.](#)
- 2 : [Ab Initio Frequently Asked Interview Questions and Answers Guide.](#)
- 3 : [Informatica Data Warehousing Frequently Asked Interview Questions and Answers Guide.](#)
- 4 : [ColdFusion Frequently Asked Interview Questions and Answers Guide.](#)
- 5 : [Cognos Data Warehouse Frequently Asked Interview Questions and Answers Guide.](#)
- 6 : [Business Objects Frequently Asked Interview Questions and Answers Guide.](#)
- 7 : [Micro Strategy Frequently Asked Interview Questions and Answers Guide.](#)
- 8 : [Warehouse Assistant Frequently Asked Interview Questions and Answers Guide.](#)
- 9 : [BO Designer Frequently Asked Interview Questions and Answers Guide.](#)
- 10 : [Warehouse DataStage Frequently Asked Interview Questions and Answers Guide.](#)



## About Global Guideline.

**Global Guideline** is a platform to develop your own skills with thousands of job interview questions and web tutorials for fresher's and experienced candidates. These interview questions and web tutorials will help you strengthen your technical skills, prepare for the interviews and quickly revise the concepts. Global Guideline invite you to unlock your potentials with thousands of [Interview Questions with Answers](#) and much more. Learn the most common technologies at Global Guideline. We will help you to explore the resources of the World Wide Web and develop your own skills from the basics to the advanced. Here you will learn anything quite easily and you will really enjoy while learning. Global Guideline will help you to become a professional and Expert, well prepared for the future.

\* This PDF was generated from <https://GlobalGuideline.com> at **November 29th, 2023**

\* If any answer or question is incorrect or inappropriate or you have correct answer or you found any problem in this document then don't hesitate feel free and [e-mail us](#) we will fix it.

You can follow us on FaceBook for latest Jobs, Updates and other interviews material.  
[www.facebook.com/InterviewQuestionsAnswers](http://www.facebook.com/InterviewQuestionsAnswers)

Follow us on Twitter for latest Jobs and interview preparation guides  
<https://twitter.com/InterviewGuide>

Best Of Luck.

Global Guideline Team  
<https://GlobalGuideline.com>  
[Info@globalguideline.com](mailto:Info@globalguideline.com)