

Warehouse DataStage Interview Questions And Answers Guide.



Global Guideline.

<https://globalguideline.com/>



Warehouse DataStage Job Interview Preparation Guide.

Question # 1

How to fix the error "OCI has fetched truncated data" in DataStage?

Answer:-

Can we use Change capture stage to get the truncated data's. Members please confirm
Dear Friend, Not only truncated data, it captures duplicates, edit, insert, unwanted data that means every changes in the before and after data

[Read More Answers.](#)

Question # 2

Explain What are other Performance tunings you have done in your last project to increase the performance of slowly running jobs?

Answer:-

Minimise the usage of Transformer (Instead of this use Copy, modify, Filter, Row Generator)
Use SQL Code while extracting the data
Handle the nulls
Minimise the warnings
Reduce the number of lookups in a job design
Use not more than 20 stages in a job
Use IPC stage between two passive stages Reduces processing time
Drop indexes before data loading and recreate after loading data into tables
Gen'll we cannot avoid no of lookups if our requirements to do lookups compulsory.
There is no limit for no of stages like 20 or 30 but we can break the job into small jobs then we use dataset Stages to store the data.
IPC Stage that is provided in Server Jobs not in Parallel Jobs
Check the write cache of Hash file. If the same hash file is used for Look up and as well as target, disable this Option.
If the hash file is used only for lookup then "enable Preload to memory". This will improve the performance. Also, check the order of execution of the routines.
Don't use more than 7 lookups in the same transformer; introduce new transformers if it exceeds 7 lookups.
Use Preload to memory option in the hash file output.
Use Write to cache in the hash file input.
Write into the error tables only after all the transformer stages.
Reduce the width of the input record - remove the columns that you would not use.
Cache the hash files you are reading from and writing into. Make sure your cache is big enough to hold the hash files.
Use ANALYZE.FILE or HASH.HELP to determine the optimal settings for your hash files.
This would also minimize overflow on the hash file.
If possible, break the input into multiple threads and run multiple instances of the job.
Staged the data coming from ODBC/OCI/DB2UDB stages or any database on the server using Hash/Sequential files for optimum performance also for data recovery in case job aborts.
Tuned the OCI stage for 'Array Size' and 'Rows per Transaction' numerical values for faster inserts, updates and selects.
Tuned the 'Project Tunables' in Administrator for better performance.
Used sorted data for Aggregator.
Sorted the data as much as possible in DB and reduced the use of DS-Sort for better performance of jobs
Removed the data not used from the source as early as possible in the job.
Worked with DB-admin to create appropriate Indexes on tables for better performance of DS queries
Converted some of the complex joins/business in DS to Stored Procedures on DS for faster execution of the jobs.
If an input file has an excessive number of rows and can be split-up then use standard logic to run jobs in parallel.
Before writing a routine or a transform, make sure that there is not the functionality required in one of the standard routines supplied in the sdk or ds utilities categories.
Constraints are generally CPU intensive and take a significant amount of time to process. This may be the case if the constraint calls routines or external macros but if it is inline code then the overhead will be minimal.
Try to have the constraints in the 'Selection' criteria of the jobs itself. This will eliminate the unnecessary records even getting in before joins are made.
Tuning should occur on a job-by-job basis.
Use the power of DBMS.
Try not to use a sort stage when you can use an ORDER BY clause in the database.
Using a constraint to filter a record set is much slower than performing a SELECT ? WHERE?.
Make every attempt to use the bulk loader for your particular database. Bulk loaders are generally faster than using ODBC or OLE.

[Read More Answers.](#)

Question # 3



Explain How is datastage 4.0 functionally different from the enterprise edition now?? what are the exact changes?

Answer:-

There are lot of Changes in DS EE. CDC Stage, Procedure Stage, Etc

[Read More Answers.](#)

Question # 4

How to pass parameters to job by using file?

Answer:-

You can do this, by passing parameters from unix file, and then calling the execution of a datastage job. the ds job has the parameters defined (which are passed by unix)

[Read More Answers.](#)

Question # 5

list out the versions of datastage Parallel , server editions and in which year they are realised?

Answer:-

Please do fish for such kind of info from the net.

[Read More Answers.](#)

Question # 6

What is DS Administrator used for - did u use it?

Answer:-

The Administrator enables you to set up DataStage users, control the purging of the Repository, and, if National Language Support (NLS) is enabled, install and manage maps and locales

[Read More Answers.](#)

Question # 7

What is a project? Specify its various components?

Answer:-

You always enter DataStage through a DataStage project. When you start a DataStage client you are prompted to connect to a project. Each project contains: DataStage jobs.

Built-in components. These are predefined components used in a job.

User-defined components. These are customized components created using the DataStage Manager or DataStage Designer

You always enter DataStage through a DataStage project. When you start a DataStage client you are prompted to connect to a project. Each project contains:

[Read More Answers.](#)

Question # 8

Explain What does a Config File in parallel extender consist of?

Answer:-

Config file consists of the following.

a) Number of Processes or Nodes.

b) Actual Disk Storage Location.

The APT_Configuration file is having the information of resource disk,node pool,and scratch information,

node information in the since it contains the how many nodes we given to run the jobs, because based on the nodes only data stage will create processors at back end while running the jobs,resource disk means this is the place where exactly jobs will be loading,scratch information will be useful whenever we using the lookups in the jobs

[Read More Answers.](#)

Question # 9

Explain How to implement type2 slowly changing dimension in datastage?

Answer:-

Slow changing dimension is a common problem in Dataware housing. For example: There exists a customer called lisa in a company ABC and she lives in New York. Later she she moved to Florida. The company must modify her address now. In general 3 ways to solve this problem

Type 1: The new record replaces the original record, no trace of the old record at all, Type 2: A new record is added into the customer dimension table. Therefore, the customer is treated essentially as two different people. Type 3: The original record is modified to reflect the changes.

In Type1 the new one will over write the existing one that means no history is maintained, History of the person where she stayed last is lost, simple to use.

In Type2 New record is added, therefore both the original and the new record Will be present, the new record will get its own primary key, Advantage of using this type2 is, Historical information is maintained But size of the dimension table grows, storage and performance can become a concern.

Type2 should only be used if it is necessary for the data warehouse to track the historical changes.

In Type3 there will be 2 columns one to indicate the original value and the other to indicate the current value. example a new column will be added which shows the original address as New york and the current address as Florida. Helps in keeping some part of the history and table size is not increased. But one problem is when the customer moves from Florida to Texas the new york information is lost. so Type 3 should only be used if the changes will only occur for a finite number of time.

[Read More Answers.](#)

Question # 10

Explain It is possible to run parallel jobs in server jobs?

Answer:-



No, It is not possible to run Parallel jobs in server jobs. But Server jobs can be executed in Parallel jobs

[Read More Answers.](#)

Question # 11

What is DS Director used for - did u use it?

Answer:-

Datastage Director is GUI to monitor, run, validate & schedule datastage server jobs.

[Read More Answers.](#)

Question # 12

Explain Dimensional modelling is again sub divided into 2 types?

Answer:-

- Star Schema - Simple & Much Faster. Denormalized form.
- Snowflake Schema - Complex with more Granularity. More normalized form.

[Read More Answers.](#)

Question # 13

How to kill the job in data stage?

Answer:-

by killing the respective processing i.d

first u have to go for manager click on job from menu bar then click on clean up resources there u will find pid with that go to administrator click on command type ds.tools there u can give the pid to kill the job

[Read More Answers.](#)

Question # 14

What is iconv and oconv functions?

Answer:-

Iconv()----converts string to internal storage format Oconv()----converts an expression to an output format

[Read More Answers.](#)

Question # 15

How to populate source files?

Answer:-

There are many ways to populate one is writting SQL statment in oracle is one way

[Read More Answers.](#)

Question # 16

Explain What happens out put of hash file is connected to transformer ..
What error it through?

Answer:-

If Hash file output is connected to transformer stage the hash file will consider as the Lookup file if there is no primary link to the same Transformer stage, if there is no primary link then this will treat as primary link itself. you can do SCD in server job by using Lookup functionality. This will not return any error code.

[Read More Answers.](#)

Question # 17

Explain Containers Usage and Types?

Answer:-

Container is a collection of stages used for the purpose of Reusability. There are 2 types of Containers.

- Local Container: Job Specific
- Shared Container: Used in any job within a project. ?

There are two types of shared container:?

- 1.Server shared container. Used in server jobs (can also be used in parallel jobs).?
- 2.Parallel shared container. Used in parallel jobs. You can also include server shared containers in parallel jobs as a way of incorporating server job functionality into a parallel stage (for example, you could use one to make a server plug-in stage available to a parallel job).regardsjagan

[Read More Answers.](#)

Question # 18

Explain What is the flow of loading data into fact & dimensional tables?

Answer:-

Here is the sequence of loading a datawarehouse.

1. The source data is first loading into the staging area, where data cleansing takes place.
2. The data from staging area is then loaded into dimensions/lookups.
- 3.Finally the Fact tables are loaded from the corresponding source tables from the staging area.



[Read More Answers.](#)

Question # 19

It is possible to access the same job two users at a time in datastage?

Answer:-

No it is not possible for the same job to be accessed by two users, because if one of the job is being used by one user then if the second one tries to open the same job then you get the error as "Job (jobname) is being accessed by another user"

[Read More Answers.](#)

Question # 20

What is job control? how can it used explain with steps?

Answer:-

JCL defines Job Control Language it is used to run more number of jobs at a time with or without using loops. steps:click on edit in the menu bar and select 'job properties' and enter the parameters as parameter prompt typeSTEP_ID STEP_ID stringSource SRC stringDSN DSN stringUsername un stringPassword pwd stringafter editing the above steps then set JCL button and select the jobs from the listbox and run the job

Job control can be acquired using job sequence in datastage 8.0.1.with or without loops.from the menu select new->sequence job and get the corresponding stages in the palette.

[Read More Answers.](#)

Question # 21

Explain What is merge and how it can be done plz explain with simple example taking 2 tables?

Answer:-

Merge is used to join two tables.It takes the Key columns sort them in Ascending or descending order.Let us consider two table i.e Emp,Dept.If we want to join these two tables we are having DeptNo as a common Key so we can give that column name as key and sort Deptno in ascending order and can join those two tables

[Read More Answers.](#)

Question # 22

What are the command line functions that import and export the DS jobs?

Answer:-

A. dsimport.exe- imports the DataStage components.

B. dsexport.exe- exports the DataStage components.

[Read More Answers.](#)

Question # 23

Explain What is the difference between LOOK UP File Stage and LookUP stage?

Answer:-

LookUP stage is used to deal on reference data set with source data .

where as LOOK UP File Stage is used to create the reference data set for the look up stage for to perform the look up operation with the source data.

[Read More Answers.](#)

Question # 24

Explain How many jobs have you created in your last project?

Answer:-

100+ jobs for every 6 months if you are in Development, if you are in testing 40 jobs for every 6 months although it need not be the same number for everybody

[Read More Answers.](#)

Question # 25

How good are you with your PL/SQL please explain?

Answer:-

You will not be writtinf pl/sql in datastage ! sql knowledge is enough ...

[Read More Answers.](#)

Question # 26

What is differences between Oracle8i/9i?

Answer:-

Multiprocessing,databases more dimesnionsal modeling

[Read More Answers.](#)

Question # 27

Explain What are the enhancements made in datastage 7.5 compare with 7.0?

Answer:-

Many new stages were introduced compared to datastage version 7.0. In server jobs we have stored procedure stage, command stage and generate report option was



there in file tab. In job sequence many stages like startloop activity, end loop activity, terminate loop activity and user variables activities were introduced. In parallel jobs surrogate key stage, stored procedure stage were introduced. For all other specifications,

[Read More Answers.](#)

Question # 28

Explain If data is partitioned in your job on key 1 and then you aggregate on key 2, what issues could arise?

Answer:-

Data will be partitioned on both the keys ! hardly it will take more for execution .

[Read More Answers.](#)

Question # 29

Explain What is the difference between Symmetrically parallel processing, Massively parallel processing?

Answer:-

Symmetric Multiprocessing (SMP) - Some Hardware resources may be shared by processor. Processor communicate via shared memory and have single operating system.

Cluster or Massively Parallel Processing (MPP) - Known as shared nothing in which each processor have exclusive access to hardware resources. Cluster systems can be physically dispersed. The processor have their own operating system and communicate via high speed network

[Read More Answers.](#)

Question # 30

Explain What are Sequencers?

Answer:-

A sequencer allows you to synchronize the control flow of multiple activities in a job sequence. It can have multiple input triggers as well as multiple output triggers. The sequencer operates in two modes: ALL mode. In this mode all of the inputs to the sequencer must be TRUE for any of the sequencer outputs to fire. ANY mode. In this mode, output triggers can be fired if any of the sequencer inputs are TRUE

[Read More Answers.](#)

Question # 31

Explain the difference between drs and odbc stage?

Answer:-

To answer your question the DRS stage should be faster than the ODBC stage as it uses native database connectivity. You will need to install and configure the required database clients on your DataStage server for it to work.

Dynamic Relational Stage was leveraged for Peoplesoft to have a job to run on any of the supported databases. It supports ODBC connections too. Read more of that in the plug-in documentation.

ODBC uses the ODBC driver for a particular database, DRS is a stage that tries to make it seamless for switching from one database to another. It uses the native connectivities for the chosen target

[Read More Answers.](#)

Question # 32

How to eliminate duplicate rows?

Answer:-

The duplicates can be eliminated by loading the corresponding data in the Hash file. Specify the columns on which you want to eliminate as the keys of hash.

You can delete duplicate records from source itself using DataStage taking option as userdefined query, instead of taking table read option. and you can use remove duplicate stage in DataStage. and using of hashfile as source also based on the hash key.

[Read More Answers.](#)

Question # 33

Explain How to handle the rejected rows in DataStage?

Answer:-

We can handle rejected rows in two ways with help of Constraints in a Transformer. 1) By Putting on the Rejected cell where we will be writing our constraints in the properties of the Transformer 2) Use REJECTED in the expression editor of the Constraint Create a hash file as a temporary storage for rejected rows. Create a link and use it as one of the output of the transformer. Apply either of the two steps above said on that Link. All the rows which are rejected by all the constraints will go to the Hash File.

[Read More Answers.](#)

Question # 34

Explain What is job control? how it is developed? explain with steps?

Answer:-

Controlling DataStage jobs through some other DataStage jobs. Ex: Consider two Jobs XXX and YYY. The Job YYY can be executed from Job XXX by using DataStage macros in Routines.

To Execute one job from other job, following steps need to be followed in Routines.

1. Attach job using DSAttachjob function.
2. Run the other job using DSRunjob function
3. Stop the job using DSStopJob function

[Read More Answers.](#)



Question # 35

Explain How to run the job in command prompt in unix?

Answer:-

Using dsjob command,
-options

dsjob -run -jobstatus projectname jobname

Dsjob-run-jobstatus-PARAM<Parameter 1>=value PARAM<Parameter 2>value.

<project name><jobname>. . If jobstatus=0 successful run,status=1and2 warnigs but success otherwise abort.

[Read More Answers.](#)

Question # 36

How to create Containers?

Answer:-

There are Two types of containers

1.Local Container

2.Shared Container

Local container is available for that particular Job only.

Where as Shared Containers can be used any where in the project.

Local container:

Step1:Select the stages required

Step2:Edit>ConstructContainer>Local

SharedContainer:

Step1:Select the stages required

Step2:Edit>ConstructContainer>Shared

Shared containers are stored in the SharedContainers branch of the Tree Structure

[Read More Answers.](#)

Question # 37

Explain how to improve the job performance?

Answer:-

by using partition technics we can improve the performance

like

hash,modules,range,random etc

[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 : [Extract, transform, load \(ETL\) Frequently Asked Interview Questions and Answers Guide.](#)
- 10 : [BO Designer 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

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