FAQ's in Talend

1. How to Deploy and Schedule the Job in Tac?

Ans:

We can deploy and schedule job by using TAC, for that what we do is, in TAC there is a job-conductor by clicking "Add" button new task will pop up in the right side of the TAC. in the execution task we need to provide the data like table name, description, job import, context, server data etc....Coming to the job import we have three options 1st option is we need to build the job in talend for that what we will do is the talend job right click we can see the option "Bult-in".

Once we select Bult-in zip file will Create in local folder that zip file we will import and we save the task.

Once task got save, we can see job conductor status is ready to deploy the job se need click deploy button once we deploy the status will change from ready to deploy to ready to run.

Go to the Job Conductor ----> Add the task ----> Click on Generate ----> Click on Deploy ---> Click on Run.

2. Difference Between TJava, TJavarow, TJavaflex?

Ans:

TJava is one of the Java components available in Talend.

The main purpose of the tJava component is whatever java code written in tJava component it will get directly execute in JVM at one time. Some functionality is not there in Talend and want to achieve that through java code we use tJava component

---tJava component will not have Edit schema

Example: - We use tJava to print massage in run console
-We can write custom java code like file java code can execute
-We can also write delete code to delete the source file from source folder.

tJavaRow:

Component has edit schema we can add or remove unwanted columns. we can be conceded the flow. We can apply functions like numeric data, for loop, if condition's etc... all this we can do by using tJavaRow

Whatever code you write in tJavaRow will execute each and every record because it is connected to flow.

tJavaflex:

tJavaflex is a combination of tJava and tJavaflex even like tjava tJavaRow, tJavaflex also has edit schema and can be connected to the flow.

In tJavaflex we have 3 sections start, main, end. start and end is not mandatory sections, in the main sections we apply functionality.

3. Difference Between Tmap and Tjoin?

Ans:

tMap wo used to join one or more tables whereas tJoin we can able to join two table at any point of time

- In tMap we can create multiple output flows with single input where as in tJoin only two output flows in inner join output and reject output.

- In tMap we can achieve inner join and left join where as in tJoin we can achieve only innerjoin tMap component is the Heart of the talend. We can do filters, Expression and Joins.

4. Difference between onSubjobOk, OnsubjobError, OncomponentOk and OnComponentError?

Ans:

OnSubjobOk - Once the previous sub job process is successful then it starts processing another sub job that as linked to the previous sub job.

OnsubjobError - Once the previous sub job process is Un-successful then it starts processing another sub job that as linked to the previous sub job.

OncomponentOk - Once the previous component process is successful then it starts processing another sub job that as linked to the previous component.

OnComponentError - Once the previous component process is Un-successful then it starts processing another sub job that as linked to the previous component.

5. Difference b/w built-in and repository?

Ans:

Built-in --- For each and every component we need to create the schema/configuration parameters manually.

Repository --- We have to create the schema/configuration parameters once in a repository and can reuse the same in all the jobs.

6. Difference b/w ETL process and ELT process?

Ans:

ETL stands for "Extract Transform load" Where we pull the data from different source system and load the data into staging area and from staging area, we apply the transtor and load in to the target.

Whereas ELT stands for " Extract load Transform" in the process data will be loaded first to the target table after loading we apply transformations on that data

7. Difference b/w When we need to use Insert/Update option and Update/Insert option?

Ans:

If we have 10000 records and out of those 100 records are insertions and remaining are updations then we should use Update/Insert option.

If we have 10000 records and out of those 9900 records are insertions and remaining are updations then we should use Insert/Update option.

8. Difference b/w tflowtoiterate and iterate to flow?

Ans:

tflowtoitarate is to iterate the cenation values for example from the data base if we get 5 records as a select result. If we want to iterate each record one time then we connect to tflowtoiterate component. for each iteration one value will be travelled forward and the next sub job will run, after that 2nd iteration 2nd value will take forward in this way 5 iteration will be done.

Whereas tIterate to flow it will take all the individual iterations and convert it into one flow and will get 5 records as 1 output.

9. Difference between Dimensions and Fact Table?

Ans:

With respect to dimensions table only we can create fact tables---imp point to mention

Dimension table will keep all attributes of an entity like say customer Dimensions for example " customer name", "customer number", "customer_ address" all these attributes which are related to the customer will keep in one table that table will call it has **Dimension table**

Where as in **Fact table** basically it contains some measurable values like amount, sales, quantity and measurable matrix that we will store in **Fact table**

10. How you can run the job parallelly?

Ans:

We can run the job parallelly by using tparallelize component all the sub job needs to connect to tparallelize component and used to run.

Otherwise, we need to bring all sub jobs to one master job without connecting with each other and in the job tab setting need to enable multithread execution then all jobs will run parallelly.

11. tSchemaComplianceCheck, tLogrow, tStatCatcher, tStatCatcher functionality?

Ans:

-tSchemaComplianceCheck - We can validate incoming schema - Length, datatype, Nullability (Mandatory Check).

-**tLogRow** - Displays the output on the console.

-**tStatCatcher** - This is the standalone component (No need to connect from any component) and provides the job audit information for each run. When we run the job

tStatcatcher Provides 2 records. 1 is at the time of the begining of the job and another is at the time of end of the job.

-**tLogCatcher** - This is the standalone component (No need to connect from any component) and provides the job audit information for each run.

When there is an error/warning in the job then this component will execute and provides only 1 record for each run.

12. Use of Joblets and Routines

Ans:

Joblets - Joblets are mainly for Re-Usable Purpose. We can create a joblet and can re-use the same across all the jobs.

Routines - Routines are user-defined functions and can be written by using Java. We can re-use the same routine across all the Jobs.

13. How to process multiple files at a time with in a single folder and multiple folders?

Ans:

Single Folder - By using tFileList component we can process multiple at a time and schema should be same for all the files.

Multiple Folders - We have to create a text file (All paths) with all the files in diff/multiple folders and pass it to tFlow to Iterate component and then tFileInputDelimited component – Target.

tFileInputDelimited -----> (Main Link) tFlowToIterate ---------> (Iterate Link) tFileInputDelimited ------> Target

14. How do you manage version management in talend?

Ans:

Right Click on the Job in the Repository and click on open another version and then click on M/m.

Whenever we want to make a modification in the existing job then we have to increase version of the job and then do the modifications.

15. How do you develop the jobs for better performance/How to increase the Heap Memory in talend?

Ans:

In a job, go to Run View and then advanced settings - enable the JVM arguments and modify it to 2GB/3GB based on the complexity of the Job.

We can enable the Cursor Size option (tOracleinput - Advanced setting - enable the cursor size - Provide the value as 10000). It fetches 10k records at a time else it will fetch record by record from database.

If we connect multiple reference/lookup files to a tMap component then it's better to use Lookup in parallel option - This will fetch the data from all reference files at a time.

16. How to pass a variable value from child to parent and parent to child in a job?

Ans:

Passing Context Values from Parent to Child - Create 2 jobs parent and child. Go to Parent Job and drag and drop the child job onto the Parent Job and then in tRunjob enable the option Transmit Whole Context. Passing Context Values from Child to Parent - Create 2 jobs parent and child.

In child job pass the context values to tBufferOutput and then go to the parent job and drag and drop the child job onto the parent job and then add the schema in tRunjob and pass it to the next component.

17. What is the use tOracleConnection and tOracleCommit?

Ans:

If we are using multiple tOracleInput/Output components in a same job it's better to use tOracleConnection. If we are not using tOracleConnection then it will open multiple connections (each component it will open 1 connection) so that it will reduce the performance. Instead of that by using tOracleConnection we can open 1 connection and can reuse the same connection across all the components.

tOracleConnection- If you use existing connection option then the data load into database will not be auto commit, so manually we need to add the oracle commit component to commit after loading.

18. Difference between UNION & UNION ALL?

Ans:

UNION - Removes duplicate rows.
UNION ALL - Does not remove the duplicate row. It returns all from all queries.

19. Difference between primary key, surrogate key?

Ans:

Both are unique identifier for each record. They are not allowed nulls and duplicate values.

Primary key we can used in dimension tables.
And surrogate key is used in fact tables

20. Difference between star schema & snowflake schema?

Ans:

-Star Schema contained the fact and dimension tables and have denormalised data and have high data redundancy.

-Snowflex Schema contains tha fact and dimension tables as well as sub dimension having normalised data and have low data redundancy and memory will be maintained well in snowflake schema.

21. How many triggers in Talend (TAC)?

Ans:

We can schedule the by using triggers in talend. There are 3 types of triggers.

Simple Triggers - By using this trigger we can run the job 1 time (Specified time)

CRON Triggers - By using this trigger we can run the jobs on daily basis (We can specify month, day, date, time).

File Triggers - By using this trigger we can run the jobs whenever the file is created/deleted/updated in a specified folder.

22. DIFFERENTIATE TRUNJOB AND JOBLETS?

Ans:

tRunJob- Here we need to first create job flow and add it into trunjob component and it perform similar functionality like joblets only but if we are using open studio, we can use trunjob component only because joblets are not available in open studio

Joblets- These are present only in enterprise edition and you can create direct job flow by creating new joblet instead of creating job and adding job like in trunjob.

23. Is left outer join is possible in tjoin?

Ans:

YES, it is possible but we need to uncheck the innerjoin (output reject) option in component tab.

24.Implicit context load and explicit context load?

Ans:

Implicit means if we create contexts in either group or file or database and apply them to whole project by drag and drop of group/using t context load (file, database) / by edit project settings.

Explicit means creating contexts explicitly for job and use them within job only.