

CGI Technical Round L1|informatics developer|3.2 years experience|Selected|

Transcribed by [TurboScribe.ai](#). [Go Unlimited](#) to remove this message.

Hi Arun. Hi. You are able to hear me right? Yes, yes.

Just give me a second, we can start, give me a second. Okay. Yeah, what is your overall experience and relevant experience in Informatica? So relevant and overall both 3.2s experience.

Okay, so current organization you are working for? LTECH Private Solution. And your notice period? So 15 days. Yeah, fine.

Can you just brief me about your current work and current project that you are working on so that we can start our process? Okay, so hi, I am Arun Gowda, I am from Bangalore and I am currently working as an Informatica developer in LTECH Private Solution Limited from November 2019. And my current project is Singtel. Singtel is actually a telecommunication company which provides services like mobile TV, IPTV, all those things.

And in this project, we have got sources and files in the form of tables and files. And we have loaded this into the stage table by using incremental loading and mapping variable. And also from stage to dimension table, we have used S3 type 1 and S3 type 2 based on the business requirement logics.

And also we have got the requirement in the form of change request and business requirement. As per that, we have developed this project. And also coming to my role and responsibilities, I used to interact with business analysts.

And also based on this requirement, I used to develop a mapping from source to stage, stage to dimension, and dimension to fact tables. And also I have developed the most complex mapping by using S3 type 2 and S3 type 1. And also I used to interact with Team Bleed about discussing about the projects and all. And also once after completing the mapping, I used to prepare unit test cases.

Okay, got it. Let me ask you a few questions in Informatica. Okay.

So, what is star schema and snowflake schema? Star schema is a fact table surrounded by dimension table. It is called star schema. In snowflake schema, it is a parent table.

One table is connected to another table. It has more physical joints and more parent table. But in star schema, it has only less physical joints and it has very less physical joints.

So, what is a mapping parameter and a mapping variable? So, mapping variable means it will change during the session runs. And that value will be stored in the repository so that we can use that value in the next session run. So, mapping parameter means it won't change.

It will remain the same constant like that. Okay. And you said in your current project, right? So, you are using an incremental load.

Yes. Okay. So, how you are doing that exactly? Like, where you are storing the variable and how exactly you are doing it in incremental project? So, in incremental loading, first we need to create a mapping variable by taking mapping variable.

So, I have created an incremental date like that. One variable I have taken and I have assigned default values like 0, 1, 0, 1 and 1900. And then in expression transformation, I have taken one flag variable.

There I have kept setMax with that variable. And then I have assigned a transaction date. So, suppose if a transaction date is to have default I have given 1900.

So, if today we have whatever the records we will load, that date will be compared. If it is greater than the incremental date which I have given, that value will be stored in the repository. So, that can be used just like next session run.

So, same thing like tomorrow if the date is like greater than today's date. So, that value will be stored in the repository and that can be used in the next session run like that. Okay.

So, mainly you are using a mapping variable for your incremental? Yes. And the column you are using is what is the frequency date something you said? Yeah. Transaction date.

Okay. Got it. So, what is SED type 1 and type 2? So, like if you want to maintain the latest version means then you have to go for SED type 1. Like historically if you want to maintain historical data means then you have to go for like SED type 2. Okay.

So, you said you have implemented SED type 2. Yes. So, can you just explain me how with the mapping how you did that? Okay. Suppose like I have taken my customers table as like source table.

And then connected to the expression transformation. So, like in expression transformation I have calculated the date based on like date entered minus 1. Then I have taken dimension table as my lookup table. From expression transformation I have taken customer ID to the lookup transformation.

So, then I have given a condition like customer ID is equal to like lookup ID. So, then I have taken like lookup employee key out to the router transformation. So, like in router

transformation I have created three groups like like and like that I have created three groups.

And then I have taken three update strategies. So, like and then connected to the dimension table. And then I have taken sequence generator to generate a primary case and connected to the dimension table.

So, like second notational of lookup ID then I have taken to the like update strategy. There like I connected to the directly dimension table since we are updating the existing insert. So, like and then have like third group I have taken notational of group notational of employee like employee key.

I connected to the update strategies. Then like there I have connected the dimension table. Then have again I have used the same sequence generator to that third dimension table.

And you said you have used three groups. So, first group I understood that you are using is not of insert. Yeah that insert for that one is insert.

And the second one is like a dd update. What is the second and third? Second is used for update and third one is also like insert existing insert. Okay.

So, for what is that condition you gave for second group and what is the condition you gave for third group? Notational of lookup employee key. And again notational of employee key. Okay.

So, for both you gave the same notational of employee. Yes. Yeah.

Okay. Existing update and existing insert for that. Right.

I got it. So, right existing insert. Yes.

You are using third group. I understood. So, for second group what is that you are connecting to target and what you are doing? So, like only like update update since start date.

Like only start date I will take and update it. Start date you will take and update it. Yeah.

Like it will be null now. Before like inserting like new record means it will be null. So, for that we need to update.

So, record coming as update. So, for that we have given a dd update like that. And we have updated that end date with date and termination whatever the date come know the data it will be updated like that.

Okay. Fine. So, you are using only that start date as a column to flag that record.

Like this is the old record like that. Yeah. Yes.

This is the old record and this is the new record. New record means it will come with the date right. Okay.

So, how you are. Yeah. I have like I am like I have written like end date is null.

Like I have written a lookup SQL overwrite query based on end date is null like that. Okay. Okay.

Okay. And you said you have used two sequence generators. Yeah.

Right. Yeah. For like insert and existing for first one like insert and existing insert.

Okay. So, for both you are using. Same.

Same. Same only. Same sequence generator only one sequence generator.

Okay. One sequence generator you have connected to. Both.

Both dimension table. Yeah. Yeah.

Yes. Yes. Fine.

Okay. So, suppose I assume that in your mapping you are using a joiner transformation. Okay.

Okay. So. Okay.

So, mainly in joiner transformation what are the joint conditions that we have. Only equal to. Okay.

And. Okay. I am using a joiner transformation.

So, obviously, I will have master and detail. Right. Yeah.

Yeah. And what are the types of joints. So, normal joints, detail joint, master outer join and fill outer join.

And if I use master outer join what is my output. Master outer join it will take less record. Less record as like a table.

Less record table to improve the performance like that. So, you are asking about like what output right. Yeah.

Yeah. Okay. So, like.

Yeah. I got it. Yeah.

I got it. So, my content only same record. And all the record from details.

Detail records. Okay. So, all records from detail matching records from.

Yeah. Yeah. Master.

Yes. Fine. And you know what is the.

What are the cache files that we have. Like in look up. Use joiner.

In joiner. Yeah. Index and data caches.

Index and data caches like index 10 it will hold like condition columns like data cache it will hold like columns only columns column ports. Okay. Fine.

So, what is difference between a connected and unconnected look up. So, connected look up it is connected to the mapping flow like unconnected look up we call it from expression transformations. So, connected in connected look up we can use both static and dynamic caches like in unconnected lookup only we can use static cache and connected lookup will return multiple value with multiple multiple values so unconnected lookup will return only value values and connected lookup will tell you which is used for like multiple value calculation and unconnected is used for single value calculation and default value will be showed in connected lookup and default value can be ignored in unconnected lookup.

Ok and what are the types of caches we have in lookup? look up so like static cache, dynamic cache and persistent cache, shared cache index cache and data cache ok so what is the difference between static cache and dynamic cache so like static cache which remains same constant that can be like reused again and again in the session like dynamic cache like if you want to update anything like that we will go for which will change during the session run, dynamic cache will be changed during the session run and it can be like used to update, insert anything like in look up cache so like in dynamic cache like if you are using dynamic cache in the sense like we will get new look up property and associate what will be given like new look up pro is equal to 0 in the sense like it won't insert or update any record 1 means it will insert, 2 means it will update like that ok so any idea on what is the use of this dynamic look up in what scenario we will be using this dynamic cache look up so like suppose in dynamic look up like if you have like same same column with multiple condition with the same row anything like that we can go for dynamic look up fine suppose assume that in your mapping I am using an aggregator transformation and I did not enable the group by port and I ran my mapping so yeah it will give it will give last record ok so what does that mean like suppose from my source I have 4 different departments of data I have a source file wherein I have 4 departments of data ok usually I need to do group by based on department but I forgot to enable that so what will be my output in this case last record will be the last record ok

and what is the use of normalizer transformation so like if you want to convert any column into row means then we have to go for normalization normalization information so like it has like a primary key like gk and gcid gk which is used as like a primary key in normalization and gcid which is used for like column occurrences ok and ok and ok I will give you one scenario ok suppose I have a flat file ok that flat file has 5 departments of data coming today ok so I need to create 5 different files based on department as it has 5 departments it need to produce me 5 output files ok ok understood by using transaction control we can do that we have like want me to explain in detail yeah what is the transaction control transformation what it will do so like it is used to control like commit and rollbacks based on that by with like commit and rollback it will control like it has some like tc continuous control like rollback rollback tc rollback and tlp tc after after commit after commit before tc rollback before tp rollback after like that it has so like by using like if you want to generate multiple output then you have to go for transaction control so every session we run it will generate a new output for like new output file name we can generate by using that so what is the union transformation so union transformation is like active and rapid transformation suppose like if you have same structure of same structure like files and tables like then we can then we can use to transfer to the target table like if you have multiple table structure multiple files with the same structure then we can use union transformation to load into the target tables so like union which acts like a union only in Oracle database so like union which has like union all in the sense it contain duplicate values also have you heard of indirect file loading yes yes suppose like we have hundreds of files we wanted to load into the one target table then we have to go for indirect file loading and first we have to do one-to-one mapping so then we have to like in infrastructure like folders there will be a source files there I have to create one file then I will add all those 100 files locations and with file name I will add it to that and that file I will give it in a source file name place so then after that I will load to the target table okay and suppose in my target table I have created one column called file name as well okay as I'm loading 10 different files right now can I get the file name also loaded to my target table yeah that we can do it by using there itself like while creating a target no there we have a file name file name we have file name like a target right side like a black mark will be there like file name by using that and also like in expression transformation by taking a variable and by assigning that file name .txt like that we can do by that way how you will do that okay suppose assume that you said in target we are enabling so I have enabled that so using expression how you will do that expression I will take a variable like the file output name is there no that I will take like a state.txt pipeline by using pipeline operator .txt so file name we have just given in the file name if you want to give file name like that suppose okay let me ask you like I have only two files okay each file has a 10 records I'm using indirect file loading I'm loading to a single target table so all the 20 records got loaded so first 10 records should have first file name next 10 records should have second file name okay into same target table right yes so how you will do that yeah we can do it by like a like while creating a table we have created like a file right file what table so target table

yeah yeah target table I have created a port called file name yeah how it will be loaded from the source from the source like same way by using indirect file number I don't know I give by you giving the location so all those locations into one directory yeah right when you do indirect file loading it will take the data okay it will take all the 20 records yeah I'm telling about the file name so first 10 records should have first file name right yeah so the obviously file name should come from source yeah then only you can connect that to the target okay so so just to correct you right source side you will have one option called the file name while importing the source definition if you enable that the file name will be loaded okay yeah yeah that in source also it's there I think I think I have seen it in target in target you will just enable so in source while importing definition will be there and have you heard of partitions yes yes what is the use of partition and what are the types of partitions suppose like in any is any session is running slowly means then we have to go for partitioning partition partitioning so like first we need to identify the partitioning like it's it in force bottleneck like target bottleneck or like so sorry mapping bottleneck or session bottleneck like that okay types of partitions same source bottleneck mapping bottleneck and mapping source target and session bottleneck have you heard of round-robin partition pass-through partition yes yes yes okay so what are those those are session partition corrector and actually I thought like performance issues like in the sense like a data partition it's a partition data that will load from node to database like that and like I don't have much experience on that so I didn't like I know the types so have you heard of pushdown optimization oh yes yes so what is that so suppose like if one even after like we will do like performance tuning this session is running slowly means then we have to go for like pushdown optimization so like a push to like it will push all the transformation logic into a scale like convert all the transformation logic into a scale segment and we will push to the source and target database okay and types of pushdown optimizations so types of full pushdown optimization so and full pushdown so source side pushdown optimization target side pushdown optimization and full pushdown optimization so yeah so what is that full pushdown means full pushdown means like it will push all the like data from all the transformation logic from to like both source and target table so but if you want to use full pushdown optimization in the sense like the both source and target should be from same database suppose I assume that you have developed one mapping okay so you are using an aggregator transformation in your mapping okay so how you will improve the performance of an aggregator transformation so by using like using filter transformation before aggregator so like filter to filter out un-aggregated values like filter out un-aggregated values so so like it will improve the performance look up at like a catch performance and also by using sorter and by checking the sorter or sorter input you can improve the performance let me ask you a few questions in order to see okay so okay what is that sequel thing that you have worked in your current product so what are types of sequels that you have worked when what are scenarios like only minus queries you have prepared or what exactly you did in sequel in sequel so by like if we join so if we join so I was just we dance yeah okay once you develop your mapping and below the

data right okay so where things are like a schema based on schema line we have given now based on that like go and open the database and like a base like slash star from the table name by by doing that like if any duplicate record slot means some by using minus operator and all these things by using that we can identify like I need to pick it or loaded like that okay so assume that you have loaded one table from your source till you set a dimension table you are using it so you have loaded your data to the dimension table okay so what all types of checks you are doing on the target table so you are just taking the count or are you performing any other validations as well between source and the loaded target table so in like you know so that's about a c-type to logics like what our analyst I mean you said in your current project you are loading the data from source table to your stage table right yeah so yeah incremental loading bus just device and table transaction just by using transaction whether all the data came to your stage table or not you should verify right yeah so that only like so like we will check based on like what are the data we actually have ordered by using like minus operator and all this thing minus operator like so from source to like from source schema like source schema to target schema minus like if any duplicate records that means it will show otherwise no no no duplicates records like that right you are saying about I mean orally you are saying so I'm asking what how you are verifying so you said you will verify whether there are any duplicates or not you said you are verifying right yeah how you are verifying that manually you are looking both the tables or are you writing any select queries for verification so manually like we are like manually we are looking not enter any skill query so like if we write any skill query like what we will get the data okay so after you load your data from stage to data warehouse right you said you are loading to dimension table yeah okay so between you are a stage table and dimension table how you are verifying whether data is loaded or not or you are just taking a sample data and you are verifying whether it is populated correctly or not or yeah in workflow monitor like yeah we have like workflow monitor there we have we can verify that whatever the records will show from source then those record will be loaded to the real life target and all it will show right right so okay you are not verifying using any SQL and all you are not verifying so what is the difference between a rank and dense rank rank in the sense like it will skip ranking like it will give like 1 2 2 4 like that so dense rank it won't say like a dense rank it won't skip any lines like that 1 2 2 3 like that it will give so what is the unique key and primary key so primary key is like a table can have like only one primary key and multiple columns can have primary key and primary key like is used to like establish relation between primary key and foreign key and unique you means like we can have a number of duplicates in unique you and we can have a number of unique you unique you also unique you also in a table okay so unique you will allow duplicates okay how many nulls it will allow okay whether it will allow any duplicate values other than nulls like can I know it will allow duplicate it will allow duplicate values what is the difference between a view and materialized view so like we use a physical table which doesn't store any value it's a virtual table doesn't store any value and material like it's a like a like a database like if I frequently update the data and infrequently insert the data into the table so like

we can use like you can perform a DML performance with DML commands and material view and we cannot perform like view DML commands and view and like view is a like material like material which is like faster we use very slower because material we can store the data in a database suppose assume that I have a table wherein I have 100 records in my table okay okay but there are duplicate records okay in my table okay so how you will identify what are the duplicate records that are there so like in Informatica or SQL SQL by using distinct can identify okay distinct will give you and by writing a SQL query okay I want to know okay let me ask you like this I have a table called employee ID employee name okay okay but I have duplicate records inserted for employee ID as well okay okay now I want to know how many duplicate records are there by using our having conditioning by

This file is longer than 30 minutes.

[Go Unlimited](#) at [TurboScribe.ai](#) to transcribe files up to 10 hours long.