**Monday, November 24, 2008**

[**SSIS QUESTIONS AND ANSWERS**](http://msbiravindranathreddy.blogspot.com/2008/11/ssis-questions-and-answers.html)

1. What is a package?  
a) A discrete executable unit of work composed of a collection of control flow and other objects, including data sources, transformations, process sequence, and rules, errors and event handling, and data destinations.  
  
2. What is a workflow in SSIS?  
A). A workflow is a set of instructions on how to execute tasks.  
(It is a set of instructions on how to execute tasks such as sessions, emails and shell commands. a workflow is created form work flow mgr.)  
  
3. What is the diff. between control flow Items and data flow Items?  
a).the control flow is the highest level control process. It allows you to manage the run-time process the run time process activities of data flow and other processes within a package.  
When we want to extract, transform and load data within a package. You add an SSIS dataflow task to the package control flow.  
  
4. What are the main components of SSIS (project-architecture)?  
A). SSIS architecture has 4 main components  
1. SSIS service  
2. SSIS runtime engine & runtime executables  
3. SSIS dataflow engine & dataflow components  
4. SSIS client’s

5.different components in SSIS package?  
a).1.control flow  
2.data flow  
3.event handler  
4.package explorer  
  
containers:::provide structure and scope to your package  
types of containers:  
i.task host container:the taskhost container services a single task.  
ii.sequence container:It can handle the flow of subset of a package and can help you drive a package into smaller more manageable process.  
uses:-

1. grouping tasks so that you can disable a part of the package that is no longer needed.  
2. narrowing the scope of the variable to a container.  
3. Managing the property of multiple tasks in one step by setting the properties of the container.  
iii.for loop container:evaluates an expression and repeats Its workflow until the expression evaluates to false.  
iv.for each loop container:defines a control flow repeatedly by using an enumerator.  
the for each loop container repearts the control flow for each member of a specified enumerator.  
  
tasks:::It provide the functionality to ur package.  
It is an individual unit of work.  
  
event handler::respond to raised events in your package.  
  
Precedence constraints::provide ordinal relationship b/w various Items in ur package.

6.how to deploy the package?  
a).to deploy the package first we need to configure some properties.  
goto project tab->package properties->we get a window,configure deployment utilIty as "true"  
mention the path as "bin/deployment"  
  
7. Connection manager:   
a).It is a bridge b/w package object and physical data. It provides logical representation of a connection at design time the properties of the connection mgr describes the physical connection that integration services creates when the package is run.  
  
8. Tell the utility to execute (run) the package?  
a) In BIDS a package that can be executed in debug mode by using the debug menu or toolbar or from solution explorer.  
In production, the package can be executed from the command line or from a Microsoft windows utility, or it can be scheduled for automated execution by using the sql server agent.  
i).goto->debug menu and select the start debugging button  
ii).press F5 key  
iii).right click the package and choose execute package.  
iv).command prompts utilities  
a).DTExecUI  
1. To open command prompt->run->type dtexecui->press enter  
2. The execute package utilIty dialog box opens.  
3. in that click execute to run the package.  
Wait until the package has executed successfully.  
b).DTExec utility  
1.open the command prompt window.  
2.command prompt window->type dtexec/followed by the DTS, SQL, or file option and the package path ,including package name.  
3. if the package encryption level is encrypt sensitive with password or encrypt all with password, use the decrypt option to provide the password.  
If no password is included, dtexec will prompt you for the password.  
4. Optionally, provide additional command-line options  
5. Press enter.  
6. Optionally, view logging and reporting information before closing the command prompt window.  
The execute package utility dialog box opens.  
7. In the execute package utility dialog box, click execute package.  
Wait until the package has executed successfully.  
v).using sql server mgmt studio to execute package  
1. In SSMS right click a package, and then click run package.  
Execute package utility opens.  
2. Execute the package as described previously.  
  
9. How can u design SCD in SSIS?  
a) Def:-SCD explains how to capture the changes over the period of time.  
This is also known as change data capture.  
type1: It keeps the most recent values in the target. It does not maintain the history.  
type2: It keeps the full history in the target database. For every update in the source a new record is inserted in the target.  
type3: It keeps current & previous information in the target.  
in-SSIS:  
-------  
type1: It can do require re-creating any aggregation that would be affected by the change.  
type2: changes can cause a serious inflation in the number of members of a dimension.  
type3: as wIth a type 1 change, type 3 change requires a dimension update, so u need to re-process  
All aggregations affected after change.  
  
10. How can u handle the errors through the help of logging in SSIS?  
a) To create an on error event handler to which you add the log error execute sql task.  
  
11. What is a logfile and how to send log file to mgr?  
a) It is especially useful when the package has been deployed to the production environment, and you can not use BIDS and VSA to debug the package.  
SSIS enables you to implement logging code through the Dts. Log method.  
When the Dts. Log method is called in the script, the SSIS engine will route the message to the log providers that are configured in the containing package.  
  
12. What is environment variable in SSIS?  
a) An environment variable configuration sets a package property equal to the value in an environment variable.  
Environmental configurations are useful for configuring properties that are dependent on the computer that is executing the package.  
  
13. about multiple configurations?  
a) It means including the xml configuration, environment variable, registry entry, parent package variable, SQL Server table, and direct and indirect configuration types.  
  
14. How to provide securIty to packages?  
a) In two ways  
1. Package encryption  
2. Password protection.  
  
15. as per error handling in T/R, which one handle the better performance? Like fail component, redirect row or ignore failure?  
a) Redirect row provides better performance for error handling.  
  
16. Staging area??  
a) It is a temporary data storage location. Where various data T/R activIties take place. A staging area is a kItchen of data warehouse.  
  
17. Task??  
a) An individual unIt of work.  
Types:-  
.....  
1. Active x script task  
2. Analysis services execute DDL task  
3. Analysis services processing task  
4. Bulk insert task \*  
5. Data flow task \*  
6. Data mining query task  
7. Execute Dts 2000 package task  
8. Execute package task \*  
9. Execute process task  
10. Execute sql task \*  
11. File system task  
12. Ftp task  
13. Message queue task  
14. Script task \*  
15. Send mail task \*   
16. Web service task  
17. Wmi data reader task  
18. Wmi event task  
19. Xml task  
  
18. Event handling & logging?  
a) You can select the t/r fails and exIts up on an error, or the bad rows can be redirected to a failed  
Data flow branch. Ignore failure, redirect row.  
Logging also improved there are more than a 12 events that can be logged for each task or package. You can enable partial logging for one task and enable much more detailed logging for billing tasks.  
Ex:-on error  
On post validate  
On progress  
On warning  
--->log file can be wrItten to usually any connection  
Sql profiler  
Text files  
Sql server  
Window event log  
Xml file  
  
19. Import & export wizard?  
a) Easiest method to move data from sources like oracle, db2, sql server.  
Right click on database name->goto task->import and export wizard  
Select the source  
Select the destination  
Query copy of tables  
Execute  
Finish  
  
20.solution explorer?  
after creating project  
project name  
-data source  
-data source views  
-packages  
-miscellaneous  
  
21. Precedence constraints?  
a) Constraints that link executable, container, and tasks wIthin the package control flow and specify condItion that determine the sequence  
And condItions for determine whether executable run.  
  
22. Data pipeline?  
a) The memory based, multIthreaded, buffered t/r process flow data through an SSIS data flow task during package execution.  
  
23. TRANSFORMATIONS??  
It is an object that generates, modifies, or passes data.  
1.AGGEGATE T/R:-It applies an aggregate function to grouped records and produces new output records from aggregated results.  
2.AUDIT T/R:-the t/r adds the value of a system variable, such as machine name or execution instance GUID to a new output column.  
3.CHARACTER MAP T/R:-this t/r makes string data changes such as changing data from lower case to upper case.  
4.CONDITIONAL SPLIT:-It separate input rows into separate output data pipelines based on the boolian expressions configured for each output.  
5.COPY COLUMN:-add a copy of column to the t/r output we can later transform the copy keeping the original for audIting personal  
6.DATA CONVERSION:-converts a columns data type to another data type.  
7.DATA MINING QUERY:-perform a data mining query against analysis services.  
8.DERIVED COLUMN:-create a new derive column calculated from expression.  
9.EXPORT COLUMN:-It allows you to export a column from the data flow to a file.  
10.FUZZY GROUPING:-perform data cleansing by finding rows that are likely duplicates.  
11.FUZZY LOOKUP:-matches and standardizes data based on fuzzy logic.  
eg:-transform the name jon to john  
12.IMPORT COLUMN:-reads the dat from a file & adds It into a dataflow.  
13.LOOKUP:-perform the lookup of data tobe used later in a transform.  
ex:-t/f to lookup a cIty based on zipcode.  
1.getting a related value from a table using a key column value  
2.update slowly changing dimension table  
3.to check whether records already exist in the table.  
14.MERGE:-merges two sorted data sets into a single data set into a single data flow.  
15.MERGE JOIN:-merges two data sets into a single dataset using a join junction.  
16.MULTI CAST:-sends a copy of two datato an addItional path in the workflow.  
17.ROW COUNT:-stores the rows count from the data flow into a variable.  
18.ROW SAMPLING:-captures the sample of data from the dataflow by using a row count of the total rows in dataflow.  
19.ROW SAMPLING:-captures the sample of the data from the data flow by using a row count of the total rows in data flow.  
20.UNION ALL:-merge multiple data sets into a single dataset.  
21.PIVOT:-converts rows into columns  
22.UNPIVOT:-converts columns into rows  
  
24. Batch?  
a) A batch is defined as group of sessions. Those are 2 types.  
1. Parallel batch processing  
2. Sequential batch processing  
  
-----For executing the package we can use "execute package utilIty"----  
  
----for deploying the package we can use "package deployment utilIty"—  
  
  
SSRS:--  
  
1. What are the main components of reporting services?  
a) Report designer, report server, report manager, report user.  
  
2. Where can u publish the report?  
a) By using report designer or publish reports on report server.  
  
3. What are the necessIty things for creating matrix report?  
a) Page, column, row, details  
  
4. for generating reports which is used like RDBMS OR CUBE?  
a) Depends on data  
  
5. What is assembly code in SSRS?  
a)  
  
6. What is Rdl file?  
a) Rdl is a Report DefinItion Language. Every report is saving wIth Rdl Extension.  
  
7. How can u create a job?  
a) Job is a group of reports.  
The reports can be processed eIther sequentially or parallel.  
  
SSAS:-  
  
1. What are the fixed measure and calculated measure?  
a) Normally we used fixed measures in SSIS mainly for calculating measures.  
Where as calculated measures uses in SSAS, while creating cube we can mention this calculated measure in the OLAP.  
  
2. What are measures?  
a) Measures are numeric data based on columns in a fact table.  
  
3. What are cubes?  
a) Cubes are data processing unIts composed of fact tables and dimensions from the data warehouse. They provided multidimensional analysis.  
  
4. What are virtual cubes?  
These are combination of one or more real cubes and require no disk space to store them. They store only definItion and not the data.  
  
  
  
  
  
  
DATAWARE HOUSE CONCEPTS:-  
  
1. Diff b/w OLTP AND OLAP?  
A)  
OLTP OLAP  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
1.transactional processing 1.query processing  
2.time sensItive 2.history oriented  
3. Operator & clerks view 3.Managers, CEOs, PM’s views  
4. organized by transaction 4.organized by subjects  
(Order, input, inventory) (product, customer)  
5.relatively smaller DB 5.large DB size  
6.volatile data 6.non-volatile  
7.stores all data 7.stores relevant data  
8. Not flexible 8.flexible  
  
2. Diff b/w star schema and snowflake?  
a) STAR SCHEMA SNOWFLAKE  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
1.centrally located fact table 1.centraly located fact table  
surrounded by de normalise surronded by the normalized  
Dimensions. dimension table.  
2.all dimensions will be link 2.all dim link wIth each other (or)  
directly wIth fact table. 1-N relationship wIth other table.  
3.It is easy to understand by 3.It is diff to understand.  
end user or tech people 4.It is diff to retrieve the data while  
4.We can easily retrieve data parsing the query against the facts n dim.  
By passing the simple queries.  
5.increase the query perform- 5.more joins.  
ance because It involve less  
Joins.  
  
What are fact tables?  
a) A fact table is a table that contains summarized numerical (facts) and historical data.  
This fact table has a foreign key-primary key relation wIth a dimension table. the fact table maintains the information in 3rd normal form.  
  
3. Types of facts?  
a)  
1. AddItive:-able to add the facts along wIth all the dimensions  
-discrete numerical measures.  
-Ex:-retail sales in $  
2. semi addItive:-snapshot taken at a point in time  
- Measure of intensIty  
-not addItive along time dimensions  
ex:-account balance, inventory balance  
3.non-addItive:-numerical measures that can't be added across any dimensions.  
-intensIty measure arranged across all dimension  
ex:-room temperatures, averages  
  
  
4. Data warehouse?  
a) A data ware house is a collection of data marts representing historical data from diff operational data sources (OLTP).  
The data from these OLTP are structured and optimized for querying and data analysis in a data warehouse.  
  
5. Data mart?  
a) A data mart is a subset of a data warehouse that can provide data for reporting and analysis on a section, unIt or a department like sales dept, hr dept.  
  
6. What is OLAP?  
a) OLAP stands for online analytical processing. It uses databases tables (fact and dimension table) to enable multi dimensional viewing, analysis and querying of large amount of data.  
  
7. What is OLTP?  
a) OLTP stands for online transactional processing. Except data warehouse databases the other databases are OLTP.  
These OLTP uses normalized schema structure.  
These OLTP databases are designed for recording the daily operations and transactions of a business.  
  
8. What are dimensions?  
Dimensions are categories by which summarized data can be viewed. For example a profIt summary fact table can be viewed by a time dimension.  
  
9. What are conformed dimension?  
a) The dimensions which are reusable and fixed in nature. Example customer, time, geography dimensions.  
  
10. Staging area?  
a) It is a temporary data storage location, where various data t/r activIties take place.  
  
11. Fact grain(granularIty)?  
a) The grain of fact is defined as the level at which the fact information is stored in a fact table.  
  
12. What is a fact less fact table?  
a) The fact table which does not contain facts is called as fact table.  
Generally when we need to combine two data marts, then one data mart will have a fact less fact table and other one wIth common fact table.  
  
13. What are measures?  
a) Measures are numeric data based on columns in a fact table.  
  
14. What are cubes?  
a) Cubes are data processing unIts composed of fact tables and dimensions from the data warehouse. They provided multidimensional analysis.  
  
15. What are virtual cubes?  
These are combination of one or more real cubes and require no disk space to store them. they store only definItion and not hte data.  
16.SCD's?  
a)  
type-I(current data)  
type-II(full historical information& Current data)  
type-III(Current data & Recent data)  
  
SQL-SERVER-2005:-  
  
1. Surrogate key?  
a)It is an artificial or synthetic key that is used as a substItute for a natural keys.  
It is just a unique identifier or number for each row that can be used for the primary key to the table.  
(It is a sequence generate key which is assigned to be a primary key in the system(table)).  
  
2.primary key?  
a)can be used to uniquely identify every row of the table.  
------unique + not null-------  
  
3.foreign key?  
a)It is a column r combination of columns that contain values that are found in primary key of some table.  
It may be null, not unique.  
  
4.composIte key?  
a)It is a primary key consisting of more than one column.  
  
4. indexes?  
a) It is an access strategy that is a way to sort and search records in the table.  
Indexes are essential to improve the speed wIth which records can be located and retrieved from a table.  
Types: - cluster index (can create only one index on table)  
Non-cluster index (can create 249 indexes on table)  
Unique index  
Composite index  
(Simple index, reverse key index, bitmap index, function index)  
  
5. View?  
a) It is used for data security reason  
To reduce the redundant data.  
  
6. Cluster?  
a) 1-many access path.  
Clusters are used to store data from diff tables in the same physical data blocks.  
  
7. Sequences?  
a) It is used to quickest way to retrieve the data.  
  
8. Cursors?  
a) Implicit cursor  
Explicit cursor  
Parameter cursor  
  
9. Triggers?  
a) Row trigger  
Statement trigger  
10. Transactions?  
Save point  
Commit & rollback.  
11. Security?  
a) Encryption  
Locking  
Level of lockings row level, page level, table level  
12.constraints?  
a)primary->  
foreign(reference)->  
check->  
unique->  
13. Diff b/w having and where?  
a) after performing 'group by' operation 'having wil again filter the records based on having condItion  
'where' is used to filter the data based on a condItion and It applies to retrive on a particular column.  
14. Joins?  
a) Join can combine the information from two tables into a single unit.  
inner join:-  
they matched the records together based on one or more common fields(matched-records only).  
outer join:-  
  
full join:-It combines the all rows on both sides of the join.  
cross join:-  
15. Union & union-all?  
a) Union:-columns, data types should be same  
Select distinct values  
Remove duplicates  
Union-all:-displays all the rows exact & duplicates.  
16. Diff b/w drop, delete & truncate?  
delete:-delete all rows at a time  
delete a single row data based on condItion.  
memory allocation will be there  
structure will be there  
truncate:-delete all rows at a time  
can't delete single row at a time  
memory allocation deleted  
table structure wil be there  
drop :- delete all rows at a time  
can't delete single row at a time  
memory allocation can be deleted  
table structure also do deleted  
----------------queries---------------  
1.Nth highest salary?  
select rownum,sal from (select rownum,sal from emp order by sal desc)  
group by rownum,sal having rownum=&N; (OR)  
select top 1 sal from (select top 10 sal from emp order by sal desc) A