SQL-ETL

Things I am doing in my current role.

Build and etl to take data from different sources and keep it in a data warehouse server.

Sql server

oracle

Od

Sp

Adhoc requirements

Monthly requirements

Bureau requirements

Monthly bureau upload automate

Daily bureau automate

Etl

Monitoring the etl packages and rectify the issues immediately.

Taking data from different sources and load it to warehouse

Load the data on an incremental basis for huge tables.

Consuming the data from flat files and load it to the table

Extracting the upi iw ow data and prepare the file to upload to rbi

Weekly/Monthly/quarterly automate

Consume the text files/excels and load the data to the database and then generate the reports. Bsr

Consume the flat file on a particular day in a month and load it to the database and then generate the report for the business people.

projects

Add/modify the existing etl packages.

Create the incremental logic to extract the data for the huge tables.

Created the package to run the procedures and paste the output in the specific folder or send the mail based on the requirement and running the ssrs reports. Consuming the excel files on a daily basis and generating the report.

Created the fully automated package which will pick the zip monthly files which will be having flat files and create the tables with monthly syntax and load to the same tables from the files. Deployed the packages in the ssis catalog and scheduled in agent to run on required frequency.

2

1

1

Need to prepare

What

Why

When

scenarios

Topics to review and prepare for interview

**SQL**– used to maintain the relational database and perform many different data

manipulation operations on the data.

SQL first appeared in 1974.

To execute queries,retrieve data from a database,inserts records in a database inserts ,updates,delete records from a database–create new databases and objects

**Data definition language (DDL):** defines the data structure of the database–CREATE, ALTER, DROP,Truncate, etc.

**Data manipulation language (DML):** manipulate existing data –UPDATE, INSERT,Delete etc.

**Data control language (DCL):** controls access–GRANT and REVOKE.

**Transaction Control Language (TCL):** to deal with the transaction operations in the database.

–COMMIT, ROLLBACK, SET TRANSACTION, SAVEPOINT, etc.

A **table** is a set of organized data in the form of rows and columns.It enables users to store and display records in the structure format.

**Fields** are the components to provide the structure for the table. It stores the same category of data in the same data type.

**A primary key** is a field or the combination of fields that uniquely identify each record in the table.it cannot be null or empty. We cannot change or delete the primary key column values. **The foreign key** is used to link one or more tables together. It is also known as the referencing key. A foreign key is specified as a key(single or collection of fields ) that is related to the primary key of another table.

**PRIMARY KEY** (ID)

**CONSTRAINT** fk\_customer **FOREIGN KEY** (Customer\_Id)

**REFERENCES** customer(ID)

**ON DELETE CASCADE**

**ON UPDATE CASCADE**

– **ON DELETE CASCADE ON UPDATE RESTRICT**;

It means if we delete any customer record from the customer table, then the related records in the contact table should also be deleted. And the ON UPDATE CASCADE will update automatically on the parent table to referenced fields in the child table(Here, it is Customer\_Id).

A **unique key** is a single or combination of fields that ensure all values stores in the column will be unique. It can accept a null value but only one null value per column is allowed. We can modify the unique key column values.

The unique key is also a unique identifier for records when the primary key is not present in the table.

Database: -A database is an organized collection of data stored electronically in a computer

system.

DBMS: -software program that primarily functions as an interface between the database and the end-user.(like a File Manager)

The database management systems can be categorized into several types. Some of the important lists are given below:

* Hierarchical databases (DBMS)
* Network databases (IDMS)
* Relational databases (RDBMS
* Object-oriented databases
* Document databases (Documentdb)
* Graph databases
* ER model databases
* NoSQL databases

RDBMS(based on a relational model): - It facilitates you to manipulate the data stored in

the tables by using relational operators. Ex: - Microsoft Access, MySQL, SQL Server, Oracle database, etc.

**Normalization** is used to minimize redundancy(causes the waste of space in the disk),inconsistency and dependency by organizing fields and tables of a database. Normalization is mainly used to add, delete or modify a field that can be made in a single table.

to remove redundancy and remove the insert, delete and update distractions. Normalization breaks the table into small partitions and then links them using different relationships to avoid the chances of redundancy

If not performing the Normalization it leads to waste of space in the disk and inconsistency . There are some rules of database normalization, which is commonly known as Normal From, and they are:

* First normal form(1NF) more than one value per record divided into separate individual records
* Second normal form(2NF): - relational must be in 1NF and relation must not contain any partial dependency. elimination of partial dependencies
* Third normal form(3NF): - if it satisfies the conditions for the second normal form and there is no transitive dependency between the non-prime attributes.to

remove the transitively dependent

* Boyce-Codd normal form(BCNF) NON PRIME ATTRIBUTES.

This inconsistency will lead to the maintenance problem and affect the ACID properties as well.

<https://www.simplilearn.com/tutorials/sql-tutorial/what-is-normalization-in-sql>

Denormalization:-to optimize the efficiency of the database infrastructure.

Normalizing results in data being split into multiple tables; denormalizing results in data from normalized tables brought in a single table.

As the data grows in the normalized tables, this query consumes more and more resources. We need to denormalize to improve performance since indexing does not give us the best performance in this case.

SQL operators

**Arithmetic operators** (mathematical operations on numerical data)**:**+-\*/% Etc **Logical operators** (evaluate the expressions and return their results in True or False)**:** ALL,ANY,UNIQUE,AND, ISNULL, EXISTS, BETWEEN, IN, LIKE, NOT, OR.

**Comparison operators**(comparisons of two values and check whether they are the same or not) **: -** (=)(!= or <>)(<)(>)(<=)(>=)(!<)(!>) etc.

**Bitwise operators**(bit manipulations between two expressions of integer type.)**:** AND (& symbol), OR (|, ^), NOT (~), etc.

**Compound operators:** perform operations on a variable before setting the variable's result to the operation's result. (+=)(-=)(\*=)(/=)(%=), etc.

**String operators**( concatenation and pattern matching of strings): + (String concatenation), += (String concatenation assignment), % (Wildcard), [] (Character(s) matches), [^] (Character(s) not to match), \_ (Wildcard match one character), etc.

view : -database object that has no values. It is a virtual table that contains a subset of data within a table

takes less space

use of a view is to implement the security mechanism CREATE VIEW view\_name AS

SELECT column\_lists FROM table\_name WHERE condition;

inline views

The inline view is a construct in Oracle SQL where you can place a query in the SQL FROM, clause, just as if the query was a table name.

What is inline view in SQL ?

* An inline view is a SELECT statement in the FROM-clause of another SELECT statement. In-line views are commonly used to simplify complex queries by removing join operations and condensing several separate queries into a single query.
  + As mentioned in the View section, a view is a virtual table that has the characteristics of a table yet does not hold any actual data.
  + In an inline view construct, instead of specifying table name(s) after the FROM keyword, the source of the data actually comes from the inline view.
  + Inline view is sometimes referred to as derived table. These two terms are used interchangeably.
  + It is a subquery that appears in the From clause of the Select statement.
  + The inline view is a construct in Oracle SQL where you can place a query in the SQL FROM, clause, just as if the query was a table name.
  + This subquery is enclosed in parenthesis and may be given an alias name. The columns selected in the subquery can be referenced in the parent query.

Syntax:

* The syntax for an inline view is,

SELECT "column\_name" FROM (Inline View);

Index : - data structure that provides a quick lookup of data in a column or columns of a table.

used to increase the performance and allow faster retrieval of records from the table Indexing reduces the number of data pages we need to visit to find a particular data page.

different types of indexes

* Unique Index

* Clustered Index
* Non-Clustered Index
* Bit-Map Index
* Normal Index
* Composite Index
* B-Tree Index
* Function-Based Index

unique index

to enforce the uniqueness of values in single or multiple columns can create more than one unique index in a single table

CREATE UNIQUE INDEX index\_name ON table\_name (index\_column1, index\_column2,...);

clustered index : - **sort and store the data rows in the table or view based on their key values**.(Like a Dictionary)

store the data information and the data itself.

There can be only one clustered index per table, because the data rows themselves can be stored in only one order.(determines the order of the table data) automatically created when the primary key is generated

non-clustered index

indexes other than PRIMARY indexes (clustered indexes) are called non-clustered indexes creates a logical ordering of data rows and uses pointers for accessing the physical data files. Therefore, reading from a clustered index is slower.

the logical order of index does not match the physical stored order of the rows on disk. It cannot be able to alter the physical order of the table and maintain the logical order of data.

can create 0 to 249 non-clustered indexes

joins : -to retrieve data from multiple tables into a meaningful result set.

* Selection is the process of identifying, accessing, and joining data into return data sets. The basic elements of selection in a SQL statement are the FROM and WHERE clauses.
  + The ANSI 89 syntax uses a cross join or comma-delimited list of tables, and places the join in the WHERE clause.

select a.id, b.id, b.address\_1 from person a, address b where a.id = b.id

* The ANSI 92 syntax uses a set of key words like INNER JOIN, LEFT JOIN, et cetera, and puts the join in an ON or USING clause, which are often referred to as subclauses.

select a.id, b.id, b.address\_1 from person a inner join address b on a.id = b.id

types of joins

* INNER JOIN: - matching records from two tables
* LEFT OUTER JOIN: -to fetch all rows from the left-hand table and common records between the specified tables.
* RIGHT OUTER JOIN : -to retrieve all rows from the right-hand table and only those rows from the other table that fulfilled the join condition.
* FULL OUTER JOIN(combination of a Right and Left Join): -Full join can return data in 3

quadrants: The rows that match, the rows that are only in the Left table and the rows that are only in the Right table. Full Joins are often used when analyzing or evaluating data.

* SELF JOIN: -used to join a table with itself,used to combine and compare the rows of the same table in the database.
* CROSS JOIN : -CROSS JOIN returns a Cartesian product, or all records joined to all records in all tables.

set operators: - to merge data from one or more tables of the same kind

SQL joins combine columns from separate tables, whereas SQL set operators combine rows

from different queries.

**UNION: -** combines two or more results and removes the duplicate rows from the tables.

**UNION ALL: -** combines two or more results and does not remove the duplicate rows from the tables.

**INTERSECT: -**returns the common records,It always retrieves unique records and arranges them in ascending order by default.

**MINUS: -**This operator returns the records from the first query, which is not found in the second query.

Constraint : - to specify the rule and regulations that allows or restricts what values/data will be stored in the table.

It ensures data accuracy and integrity inside the table.

It enforces us to store valid data and prevents us from storing irrelevant data.

Ex: - NOT NULL, PRIMARY KEY, FOREIGN KEY, AUTO\_INCREMENT, UNIQUE KEY, etc.

The following constraints are commonly used in SQL:

* [NOT NULL](https://www.w3schools.com/sql/sql_notnull.asp) - Ensures that a column cannot have a NULL value
* [UNIQUE](https://www.w3schools.com/sql/sql_unique.asp) - Ensures that all values in a column are different
* [PRIMARY KEY](https://www.w3schools.com/sql/sql_primarykey.asp) - A combination of NOT NULL and UNIQUE. Uniquely identifies each row in a table
* [FOREIGN KEY](https://www.w3schools.com/sql/sql_foreignkey.asp) - Prevents actions that would destroy links between tables
* [CHECK](https://www.w3schools.com/sql/sql_check.asp) - Ensures that the values in a column satisfies a specific condition
* [DEFAULT](https://www.w3schools.com/sql/sql_default.asp) - Sets a default value for a column if no value is specified
* [CREATE INDEX](https://www.w3schools.com/sql/sql_create_index.asp) - Used to create and retrieve data from the database very quickly

DELETE—t**o delete a row,slower because it maintains the log,can roll back,and occupies more space.**

**TRUNCATE**

**What are the differences between DELETE and TRUNCATE?**

|  |  |
| --- | --- |
| **DELETE COMMAND** | **TRUNCATE COMMAND** |
| **DELETE command is a DML** | **TRUNCATE command is a DDL command.** |
| **command.** |  |
| **It is executed using a row lock,** | **It is executed using a table lock, and the whole table** |
| **each row in the table is locked for** | **is locked to remove all records.** |
| **deletion.** |  |
| **WHERE clause can be used with** | **WHERE clause can’t be used with TRUNCATE.** |
| **DELETE to filter and delete specific** |  |
| **data.** |  |
| **It is used to remove rows from a** | **It removes all rows from a table.** |
| **table based on WHERE condition.** |  |
| **It maintains the log, so it’s slower** | **It is faster performance wise because it has minimal** |
| **than TRUNCATE.** | **logging in the transaction log.** |
| **It removes rows one at a time and** | **It removes the data by deallocating the data pages** |
| **records an entry in the transaction** | **that are used to store the table data. It records only** |
| **log for each deleted row.** | **the page deallocations in the transaction log.** |
| **It uses more transaction space** | **It uses less transaction space that the DELETE** |
| **than TRUNCATE statement.** | **command.** |
| **It can be used with indexed views.** | **It cannot be used with indexed views.** |
| **DELETE command active** | **​ whereas TRUNCATE does not.** |
| **triggers** |  |
| TRansactions in sql: - |  |

A SQL transaction is **a grouping of one or more SQL statements that interact with a database**. A transaction in its entirety can commit to a database as a singlelogical unit or rollback (become undone) as a single logical unit.

A transaction is **the propagation of one or more changes to the database**.

**Transaction is a set of operations performed in a logical sequence. It is executed as a whole, if any statement in the transaction fails, the whole transaction is marked as failed and not committed to the database.**

ACID property**:-contribute to the ability of a transaction to ensure data integrity.**

* Atomicity: This property ensures that the transaction(insert, update, delete) is completed in all-or-nothing way.

If one part of the transaction fails, the entire transaction fails, and the database state is left unchanged.

Its main features are COMMIT, ROLLBACK, and AUTO-COMMIT.

* Consistency: This ensures that updates made to the database are valid and follows rules and restrictions.database changes state only when a transaction will be

committed successfully.

transaction is completed according to the defined rules, constraints, cascades, and triggers

* Isolation: Every transaction is individual, and One transaction can’t access the result of other transactions until the transaction completes. Or, you can’t perform the same operation using multiple transactions at the same time.
* Durability: This property ensures that the committed transactions are stored permanently in the database.

Once the transaction is completed, then the changes it has made to the database will be permanent. Even if there is a system failure or any abnormal changes also, this property will safeguard the committed data.

Transaction Control

The following commands are used to control transactions.

* COMMIT − to save the changes.
* ROLLBACK − to roll back the changes.
* SAVEPOINT − creates points within the groups of transactions in which to ROLLBACK.
* SET TRANSACTION − Place a name on a transaction.

The RELEASE SAVEPOINT command is used to remove a SAVEPOINT that you have created.

RELEASE SAVEPOINT SAVEPOINT\_NAME;

The SET TRANSACTION command can be used to initiate a database transaction.

This command is used to specify characteristics for the transaction that follows. For

example, you can specify a transaction to be read only or read and write.

The syntax for a SET TRANSACTION command is as follows.

SET TRANSACTION [ READ WRITE | READ ONLY ];

functions : - contains a **set of SQL statements to perform a specific task**.

to replicate the common task easily

**The following are the rules for creating SQL Server functions:**

* A function must have a name, and the name cannot begin with a special character such as @, $, #, or other similar characters.
* We can use a function anywhere such as AVG, COUNT, SUM, MIN, DATE, and other functions with the SELECT query in SQL.
* Whenever a function is called, it compiles.
* Functions must return a value or result.
* Functions use only input parameters.
* We cannot use TRY and CATCH statements in functions.

Types of Functions

SQL Server categorizes the functions into two types:

* System Functions
* User-Defined Functions(**created by the user)** System Functions(**built-in functions)**
* String Functions (LEN, SUBSTRING, REPLACE, CONCAT, TRIM)
* Date and Time Functions (datetime, datetime2, smalldatetime)
* Aggregate Functions (COUNT, MAX, MIN, SUM, AVG,FIRST(),LAST() )
* Mathematical Functions (ABS, POWER, PI, EXP, LOG)
* Ranking Functions (RANK, DENSE\_RANK, ROW\_NUMBER, NTILE)
* case manipulation functions: - LOWER ,UPPER ,INITCAP
* Character manipulation

functions:-Concat,**SUBSTR,LENGTH,INSTR,LPAD,RPAD,TRIM,REPLACE**

The MOD function returns the remainder in a division operation.

LOCATE() - Used to fnd the index of the character in a string. Syntax - LOCATE(character,ColumnName)

COALESCE(NULL, NULL, NULL) will return NULL value as no non-NULL value is encountered in the parameters list.

Ques.49. What are orphan records?

Ans. Orphan records are the records having foreign key to a parent record which doesn't exist or got deleted.

Ques.50. How can we remove orphan records from a table?

Ans. In order to remove orphan records from the database we need to create a join on the parent and child tables and then remove the rows from the child table where id IS NULL.

DELETE PT

**user-defined functions mainly into two types:**

1. **Scalar Functions : - accepts parameters, either single or multiple and returns a single value.**

**CREATE FUNCTION schema\_name.function\_name (parameter\_list) RETURNS data\_type AS**

**BEGIN statements RETURN value**

**END**

1. **Table-Valued Functions: -user-defined function that returns data of a table type(can use same as table)**

**Table-valued functions are commonly used as parameterized views. Table-valued functions are more flexible than stored procedures because they can be used anywhere tables are used.**

1. **Inline Table-Valued Functions: -This UDF function returns a table variable based on the action performed by the function.**

**--It creates a table-valued function to get employees CREATE FUNCTION fudf\_GetEmployee() RETURNS TABLE**

**AS**

**RETURN (SELECT \* FROM Employee)**

**We can use like SELECT \* FROM udf\_GetEmployee(); .**

**2. Multi-statement table-valued functions (MSTVF): -returns a table variable**

**It can contain single or multiple statements to produce the result, and it is also a function that returns the result of multiple statements in a tabular form.**

**CREATE FUNCTION MULTIVALUED()**

**RETURNS @Employee TABLE**

**(id INT, emp\_name VARCHAR(50), salary INT) AS BEGIN**

**INSERT INTO @Employee**

**SELECT E.id, E.emp\_name, E.salary FROM Employee E;**

**UPDATE @Employee SET emp\_name = 'Graeme Smith' WHERE id = 3;**

**RETURN**

**END**

**We can use this as SELECT \* FROM MULTIVALUED();**

SQL Injection: - type of vulnerability in website and web app code that allows attackers to control back-end operations and access, retrieve, and destroy sensitive data from databases.

**Cursor** is a Temporary Memory or Temporary Work Station. It is Allocated by Database Server at the Time of Performing DML(Data Manipulation Language) operations on Table by User. Cursors are used to store Database Tables. There are 2 types of Cursors: Implicit Cursors, and Explicit Cursors. These are explained as following below.

1. **Implicit Cursors:**

Implicit Cursors are also known as Default Cursors of SQL SERVER. These Cursors are allocated by SQL SERVER when the user performs DML operations.

1. **Explicit Cursors :**

Explicit Cursors are Created by Users whenever the user requires them. Explicit Cursors are used for Fetching data from Table in Row-By-Row Manner.

**Fetch Data from cursor.**

There are total 6 methods to access data from cursor. They are as follows :

**FIRST** is used to fetch only the first row from cursor table.

**LAST** is used to fetch only last row from cursor table.

**NEXT** is used to fetch data in forward direction from cursor table. **PRIOR** is used to fetch data in backward direction from cursor table. **ABSOLUTE n** is used to fetch the exact nth row from cursor table. **RELATIVE n** is used to fetch the data in incremental way as well as decremental way.

**Syntax :** FETCH NEXT/FIRST/LAST/PRIOR/ABSOLUTE n/RELATIVE n FROM cursor\_name

**Cursor: -allows for the traversal of records in a database**

Working with SQL Cursor:

1. DECLARE a cursor after any variable declaration. The cursor declaration must always be associated with a SELECT Statement.
2. Open cursor to initialize the result set. The OPEN statement must be called before fetching rows from the result set.
3. FETCH statement to retrieve and move to the next row in the result set.

1. Call the CLOSE statement to deactivate the cursor.
2. Finally use the DEALLOCATE statement to delete the cursor definition and release the associated resources.

**DECLARE** @name VARCHAR(50) /\* Declare All Required Variables \*/

**DECLARE** db\_cursor **CURSOR FOR**/\* Declare Cursor Name\*/ **SELECT** name

**FROM** myDB.students

**WHERE** parent\_name **IN** ('Sara', 'Ansh')

**OPEN** db\_cursor /\* Open cursor and Fetch data into @name \*/

**FETCH** next

**FROM** db\_cursor

**INTO** @name

**CLOSE** db\_cursor /\* Close the cursor and deallocate the resources \*/ **DEALLOCATE** db\_cursor

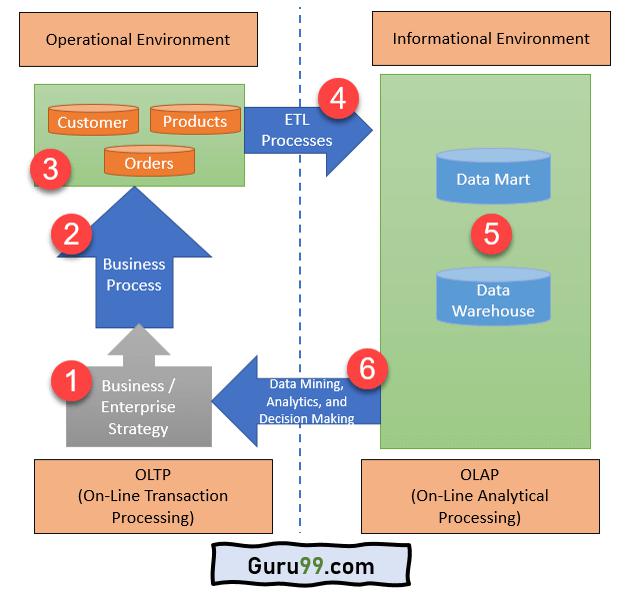
**OLTP: - (Online Transaction Processing): -** focused on query processing, maintaining data integrity in multi-access environments as well as effectiveness that is measured by the total number of transactions per second. An example of the OLTP system is the ATM center.

**OLAP : -**stands for Online Analytical Processing,

It is characterized by a large volume of data.

OLAP uses the data warehouse.

It is designed for analysis of business measures by category and attributes.



**Collation : -Collation refers to a set of rules that determine how data is sorted and compared.**

* Case sensitivity: A and a are treated differently.
* Accent sensitivity: a and á are treated differently.
* Kana sensitivity: Japanese kana characters Hiragana and Katakana are treated differently.
* Width sensitivity: Same character represented in single-byte (half-width) and double-byte (full-width) are treated differently

**Stored Procedure**: - A stored procedure is **a prepared SQL code that you can save, so the code can be reused over and over again.**

A stored procedure is a **group of one or more pre-compiled SQL statements**

It also provides a sense of security and functionality as users who can't access the data directly can be granted access via stored procedures.

**CREATE PROCEDURE** [schema\_name].procedure\_name

@parameter\_name data\_type,

....

parameter\_name data\_type

**AS**

**BEGIN**

* SQL statements
* SELECT, INSERT, UPDATE, or DELETE statement

**END**

Features of Stored Procedures in SQL Server

The following are the features of stored procedure in SQL Server:

* **Reduced Trafc:** A stored procedure reduces network trafc between the application and the database server, resulting in increased performance. It is because instead of sending several [SQL](https://www.javatpoint.com/sql-tutorial) statements, the application only needs to send the name of the stored procedure and its parameters.
* **Stronger Security:** The procedure is always secure because it manages which processes and activities we can perform. It removes the need for permissions to be granted at the database object level and simplifes the security layers.
* **Reusable:** Stored procedures are reusable. It reduces code inconsistency, prevents unnecessary rewrites of the same code, and makes the code transparent to all applications or users.
* **Easy Maintenance:** The procedures are easier to maintain without restarting or deploying the application.
* **Improved Performance:** Stored Procedure increases the application performance. Once we create the stored procedures and compile them the frst time, it creates an execution plan reused for subsequent executions. The procedure is usually processed quicker because the query processor does not have to create a new plan.

Types of Stored Procedures

SQL Server categorizes the stored procedures mainly in two types:

1. User-defned Stored Procedures
2. System Stored Procedures

The server's administrative tasks depend primarily on system stored procedures.

**5.What is the raw datatype?**

**Answer:**

Raw datatype is used to store values in binary data format.

There are 2 types of RAW datatype:

1.Raw

2.Long Raw.

Long raw datatypes are used to store graphics,sound documents.

Raw datatype is variable length data type like varchar2 but basically it only stores data in 1 ‘s and 0’s means binary data format.

Ques.41. What are locks in SQL?

Ans. Locks in SQL are used for maintaining database integrity in case of concurrent execution of the same piece of data.

Ques.42. What are the different types of locks in the database?

Ans. The different types of locks in database are-

Shared locks - Allows data to be read-only(Select operations), prevents the data to be updated when in shared lock.

Update locks - Applied to resources that can be updated. There can be only one update lock on a data at a time.

Exclusive locks - Used to lock data being modifed(INSERT, UPDATE, or DELETE) by one transaction thus ensuring that multiple updates cannot be made to the same resource at the same time.

Intent locks - A notifcation mechanism using which a transaction conveys that intends to acquire lock on data.

Schema locks- Used for operations when schema or structure of the database is required to be updated.

Bulk Update locks - Used in case of bulk operations when the TABLOCK hint is used.

Triggers in SQL Server

The trigger is a database object similar to a stored procedure that is executed automatically when an event occurs in a database. There are different kinds of events that can activate a trigger like inserting or deleting rows in a table, a user logging into a database server instance, an update to a table column, a table being created, altered, or dropped, etc.

There are three types of triggers in SQL Server

* [DML triggers](https://www.tutorialsteacher.com/sqlserver/triggers#dml-triggers) are automatically fired when an INSERT, UPDATE or DELETE event occurs on a table.
* [DDL triggers](https://www.tutorialsteacher.com/sqlserver/ddl-triggers) are automatically invoked when a CREATE, ALTER, or DROP event occurs in a database. It is fired in response to a server scoped or database scoped event.
* [Logon trigger](https://www.tutorialsteacher.com/sqlserver/logon-triggers) is invoked when a LOGON event is raised when a user session is established.

INSTEAD OF Triggers

AFTER Triggers

FOR Triggers

CREATE TRIGGER [schema\_name.]trigger\_name

ON table\_name

{FOR | AFTER | INSTEAD OF} {[INSERT] [,] [UPDATE] [,] [DELETE]}

AS

{sql\_statements}

What's new in 2017 sql version: -

How will you handle performance issues in sql server

1. Query Taking Long time than usual processing
2. Blockage–wait type—

<https://www.sqlshack.com/troubleshooting-using-wait-stats-in-sql-server/>

Writelog : - Write log means there is a delay in writing to the log file.

in the given database both data and log files in the same drive. In the case of SQL Server, the first data is written to the log file and then the data is written to the data file via CHECKPOINT operation. When both files are in the same drive, there will be contention in writing as indicated in the above result. The obvious action was to move the database log file to a different drive. After moving the log file to a different drive, the problem was solved.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Wait Types | Reason | Actions | | | Avoid |
| CXPACKET | Parallel processing | 1. Most of the time this | | | Do not avoid parallelism. |
|  |  | is due to not the | | |  |
|  |  | existence of proper | | |  |
|  |  | indexes. | | |  |
|  |  | 2.Examine the query | | |  |
|  |  | plan of the query. | | |  |
|  |  | [3.Read more details at](https://www.sqlshack.com/how-to-avoid-cxpackets/) | |  |  |
|  |  | [CXPACKT.](https://www.sqlshack.com/how-to-avoid-cxpackets/) |  | |  |
| ASYNC\_NE | SQL Server is | Re-write the application | | | By looking at the name, |
| TWORK\_IO | waiting for a client to | code | | | you might think that it is |
|  | acknowledge receipt |  |  |  | something to do with the |
|  | of sent data. |  |  |  | network. |

BACKUPBU During the database

FFER backup, these types

of wait types will

occur.

BACKUPIO

BACKUPTH

READ

PAGEIOLAT Waiting for a data file

CH\_SH page to be read from

disk to memory.

PAGEIOLAT

CH\_EX

Most likely these types of wait types indicate, slowness of in backup disk.

Need to analyze what tables are having the issue.

If it is a large database, you need to leave this until the backup is completed.

Nothing to do with the IO subsystem.

1. Does the database have enough available free space?
2. Database may face out of space error
3. Implement indexes that provide a benefit to your queries.
4. Place your tempdb files on dedicated high-speed disks
5. Make TempTB have the number of files equal to the CPU up

to 8, make sure they are sized the same.

1. Grow your tempdb data files to a desired size
2. Does the database have enough available free space?
3. Are the statistics current?
4. Make sure you’re running index maintenance frequently.
5. Check your most expensive queries and stored procedures.
6. SQL Server Health Check
7. If these tips don’t help you or if you don’t have the time to investigate the underlying issue, it may be worth considering a SQL Server Health Check.
8. **Review Resource Waits in the Activity Monitor**
9. Keep in mind that anything less than 1000 ms (Recent Wait Time) is considered normal.

***Ssis: - SQL Server Integration Services***

introduced with SQL Server 2005 and it replaced the **DTS (Data Transformation Services)** in SQL Server 2000.

SSIS used to store and manage their data in warehouses, and further extract and load it whenever required.

SSIS package: - organized set of connections such as control events, data flow elements, parameters, event handlers, variables and configurations.

SSIS packages are created in **Business Intelligence Development Studio (also called BIDS in short)**.

used to automate the maintenance of SQL Server databases, update multidimensional cube data, and perform other functions.

components/parts of the SSIS package

* Control Flow
* Data Flow
* Package Explorer
* Event handler

Data Flow: - the flow of data pumping from the corresponding source to a target destination.

There are three components of data flow, such as **Source, Transformation,** and **Destination**.

* Data Flow is the subset of control flow
* There will be only one control flow while multiple dataflow can exist.

**Control Flow:** - to link and program different tasks using the **logical connectors** and control the elements

In SSIS, there are mainly three following types of control flow elements:

* Tasks that provide **functionalities** (such as file system tasks,create backups, connect to FTP, execute scripts or SQL tasks, etc.) in packages.
* Containers that provide **structures** (such as sequential, for loop, for each loop).
* **Precedence Constraints** that join tasks, containers, and executable in a sequential control flow.

Data Transformation: - to manage and transfer the extracted data to a separate file we have selected.

**the components in the data flow of a package that aggregate, merge, distribute, and modify data**.

**Transformations: -**

1. SSIS Aggregate Transformation
2. SSIS Data Conversion Transformation
3. SSIS Copy Column Transformation
4. SSIS Conditional Split Transformation
5. SSIS Data Conversion Transformation
6. SSIS Derived Column Transformation
7. SSIS Lookup Transformation

1. SSIS Union All Transformation
2. SSIS Merge Transformation
3. SSIS Merge Join Transformation
4. SSIS Multicast Transformation
   1. SSIS Audit Transformation
   2. SSIS Cache Transform Transformation
   3. SSIS Character Map Transformation
   4. SSIS Data Mining Query Transformation
   5. SSIS DQS Cleansing Transformation
   6. SSIS Export Column Transformation
   7. SSIS Fuzzy Grouping Transformation
   8. SSIS Fuzzy LookUp Transformation
   9. SSIS Import Column Transformation
   10. SSIS OLEDB Command Transformation
   11. SSIS Percentage Sampling Transformation
   12. SSIS PIVOT Transformation
   13. SSIS Row Count Transformation
   14. SSIS Row Sampling Transformation
   15. SSIS Script Component Transformation
   16. SSIS Slowly Changing Dimension Transformation
   17. SSIS Sort Transformation
   18. SSIS Term Extraction Transformation
   19. SSIS Term LookUp Transformation
   20. SSIS UnPivot Transformation

In SSIS, transformations are available in two main categories–Synchronous and Asynchronous. During ETL design it’s recommended to use all Synchronous transformation components.

Synchronous are components like the Conditional Split or Derived Column Transformation where rows flow into memory buffers in the transformation and the same buffers come out. No rows are held and characteristically these transformations perform very quickly with marginal impact to Data Flow.

Output is synchronous with the input, it occurs at the same time and to process the given row does not need the information about the other row.

Ex : - Derived column,derived column,data conversion

Asynchronous transformation has two types, fully blocking and partial blocking. Partial blocking transformation is that transformation which creates new memory buffers for the output of the transformation than what comes into the transformation, like Union All Transformation; fully blocking transformations also require a new memory buffer similar to

partial blocking. Asynchronous transformations additionally cause a full block of the data like Sort and Aggregate transformations.

Each row cannot be processed individually; there are dependencies of one record with another record.

|  |  |  |
| --- | --- | --- |
| Ex: - Sort ,aggregate |  |  |
|  |  |  |
| Synchronous | Asynchronous Transformations | |
| Transformations |  |  |
|  |  |  |
|  | Partial | Fully |
|  | blocking | blocking |
|  |  |  |
| Audit | Data | Aggregate |
|  | Mining |  |
|  | Query |  |
|  |  |  |
| Character Map | Merge | Fuzzy |
|  |  | Grouping |
|  |  |  |
| Conditional Split | Merge Join | Fuzzy |
|  |  | Lookup |
|  |  |  |
| Copy Column | Pivot | Row |
|  |  | Sampling |
|  |  |  |
| Data Conversion | Term | Sort |
|  | Lookup |  |
|  |  |  |
| Derived Column | Union All | Term |
|  |  | Extraction |
|  |  |  |
| Export Column | Unpivot |  |
|  |  |  |

Import Column

Lookup

Multicast

OLE DB Command

Percent Sampling

Row Count

Script Component

Slowly Changing

Dimension

Partially\ Semi

Blocking Non-Blocking

Blocking

|  |  |
| --- | --- |
| Need to read all the |  |
| inputs before they |  |
| Need not read | Need not read |
| produce some output. |  |
| all inputs | all inputs |
| Clearly, it consumes |  |
| time |  |
|  |  |

|  |  |  |
| --- | --- | --- |
|  | It uses the |  |
|  | existing buffer. |  |
|  | It creates a new |  |
| It creates a new buffer | This introduces |  |
|  | buffer for the |  |
| for the output | huge |  |
|  | output |  |
|  | performance |  |
|  | benefits |  |
|  |  |  |
| These transformations | Number of |  |
|  |  |
| may create additional | Number of |  |
|  | output records |  |
| records i.e. the | output records |  |
|  | may or may not |  |
| number of input | are same as |  |
|  | match the |  |
| records might be | number of input |  |
|  | number of input |  |
| different from number | records. These |  |
|  | records. They |  |
| of output records. | transformations |  |
|  | are also |  |
| Hence, they are | are synchronous |  |
|  | asynchronous in |  |
| asynchronous in | in nature |  |
| nature. | nature |  |
|  |  |
|  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| New thread is | New thread | No additional |  |
| introduced in Data | might introduce |  |
| thread is created |  |
| flow | in data flow |  |
|  |  |
|  |  |  |  |
|  |  | It has the |  |
|  | Less impact on | minimum impact |  |
| Available resources | available | to available |  |
| resources than | resources out of |  |
| are impacted heavily |  |
| Blocking | all the |  |
|  |  |
|  | transformations | transformation |  |
|  |  | types |  |
|  |  |  |  |
|  |  | Ex: Data |  |
| Ex: Sort, Aggregate | Ex: Merge, | Conversion, |  |
| Merge Join | Derive Column, |  |
|  |  |
|  |  | Look-up |  |

Transactions in SSIS

Transaction support is built into SSIS. The TransactionOption property exists at the package level, container level (e.g. For Loop, Foreach Loop, Sequence, etc.), as well as just about any Control Flow task (e.g. Execute SQL task, Data Flow task, etc.). TransactionOption can be set to one of the following:

* Required - if a transaction exists join it else start a new one
* Supported - It does not start a new transaction,if a transaction exists join it (this is the default)
* NotSupported - do not join an existing transaction

Transaction Isolation Levels

The transaction isolation level determines the duration that locks are held. We'll use SQL Server as an example. The following transaction isolation levels are available in SQL Server:

* READ UNCOMMITTED - reads do not acquire share locks and they don't wait on locks. This is often referred to as a dirty read because you can read modifed data that hasn't been committed yet and it could get rolled back after you read it.
* READ COMMITTED - reads acquire share locks and wait on any data modifed by a transaction in process. This is the SQL Server default.
* REPEATABLE READ - same as READ COMMITTED but in addition share locks are retained on rows read for the duration of the transaction. In other words any row that is read cannot be modifed by another connection until the transaction commits or rolls back.
* SERIALIZABLE - same as REPEATABLE READ but in addition no other connection can insert rows if the new rows would appear in a SELECT

statement already issued. In other words if you issue a select statement in a transaction using the SERIALIZABLE isolation level you will get the same exact result set if you issue the select statement again within the same transaction.

SQL Server 2005 added two new options:

* A variation on READ COMMITTED where you set

READ\_COMMITTED\_SNAPHOT ON at the database level and any transaction that uses the READ COMMITTED isolation level will not acquire share locks and will not wait on any locks. Rather, you will get the committed version of all rows at the time the SELECT statement begins.

* A new isolation level called SNAPSHOT where you set

ALLOW\_SNAPSHOT\_ISOLATION ON at the database level and any transaction that explicitly sets the transaction isolation level to snapshot will not acquire share locks and will not wait on any locks. Rather, you will get the committed version of all rows at the time the transaction begins.

* Chaos
* Unspecified

SSIS Catalog: - used to store all the deployed packages,used to increase the security of the stored packages

Task: - carry out an individual component of work divided into the following two categories:

* Database Maintenance Tasks
* Control Flow Tasks

Bulk Insert Task: - used to upload vast amounts of data from flat files to SQL Server. In this process, only OLE DB connections to the destination database are supported. Container: - the set of tasks linked logically

four types of containers used in SSIS:

* **Task Host container-** Only one task is placed inside the container. This is the default container.
* **Sequence Container –** This container can be defned as a subset of package control fow.
* **For loop container –** Allows looping based on condition. Runs a control fow till the condition is met.
* **for each loop container –** Loop through container-based on the enumerator.

**types of connection or files that support SSIS**

* ODBC
* OLEDB
* .net SQLClient
* Flat File
* Excel
* XML

Precedence Constraint : - commands that instruct the system to follow the specific order to execute tasks. **Success,Failure,Completion.**

Breakpoint: - allows developers to pause the package's execution in the business intelligence system.o suddenly stop the execution at any desired point and enables the developers to re-review the status of variables, data or the entire package.

A point in the processing of a program that the programmer wants to observe more closely by stopping the program and examining the contents of variables, buffers and memory

checkpoint : - saving point that helps developers to restart or load the project from any specific point.

used to restart the project from the point of failure instead of loading the entire project again.

checkpoint data is not saved for Loop containers CheckpointUsage property set to Always or IfExists

CHECKPOINT in the SSIS package works on the control flow level. We cannot configure it to execute the individual task at the data flow level.

Check point data is not saved for each loop and for loop containers.

Lookup Cache Modes

**Full Cache Mode**: it fully loads the lookup table (reference dataset) in memory before executing the lookup transformation. It is the only available mode using a cache connection manager.

**Partial Cache Mode**: the row is cached into the SSIS lookup cache only in the case when there is a subsequent match. Once the cache gets full, SSIS automatically starts removing

existing rows based on the match and usage stats. After that, new matching rows are loaded into the lookup cache.

**No Cache Mode**:the lookup data is generated and loaded during the package execution. No data is stored within the cache memory.

logging SSIS

**text files, XML files, SQL Server, SQL Server Profiler, Windows Event Log**, etc.

most critical errors

**Data Connection Errors**:

**Data Transformation Errors**:

**Expression Evaluation Errors**:

Workflow : - sequential set of instructions or commands that instruct the program executor to follow specific flow when executing tasks and containers within an SSIS package. Data profiling task : - analysis is done to better understand the source data's state in terms of data cleaning, data pattern identification, numbers or nulls in data, etc.

doesn't help while developing standard recurring ETL packages.

DelayValidation Property:

DelayValidation Property is available on Task level, Connection Manager, Container and on Package level. By default the value of this property is set to false that means that when the package start execution, It validates all the Tasks, Containers, Connection Managers and objects( Tables,Views, Stored Procedures etc.) used by them. If any object such as table or destination file etc. is not available then Package validation fails and Package stop execution.

By setting this property to True, We enforce our SSIS Package not to validate that Task, Connection Manager or entire Package at start but validate at run time.

Real Time Examples for Using DelayValidation Property in SSIS Package:

**Example 1: Make Use of Temp Table in SSIS Package**

**Example 2: Create Excel File with DateTime**

**Example 3: How to Create Multiple Files Dynamically From a SQL Server Table**

**Top 10 Frequently Asked SSIS Interview Questions**

1. [What are the Differences between SSIS and Informatica?](https://mindmajix.com/ssis-interview-questions#ssis-and-informatica)

s:

|  |  |  |
| --- | --- | --- |
| Feature | SSIS | Informatica |
|  |  |  |
| Security | Secure to use | Secure to use |
|  |  |  |
| Administration & Maintenance | Easy to use | Difficult to maintain |
|  |  |  |
| Ease of Use | Easy to implement SSIS | Moderate |
|  |  |  |
| Performance | Performs well for Medium SQL | Performs well for Medium SQL |
|  | server environment | server environment |
|  |  |  |
| Productivity | Moderate | High while implementing large |
|  |  | applications |
|  |  |  |
| Product Maturity | Younger to Informatica | Elder than SSIS |
|  |  |  |
| Cost | No cost | Expensive |
|  |  |  |

1. [What is the Difference between Execute TSQL Task and Execute SQL Task?](https://mindmajix.com/ssis-interview-questions#tsql-task) Execute the TSQL Task:
   * Pros: Takes less memory, faster performance
   * Cons: Output into variable not supported, Only supports ADO.net connection Execute SQL Task:
   * Pros: Support output into variables and multiple types of connection, parameterized query possible.
   * Cons: Takes more memory, slower performance compared to the TSQL task.
2. [What is the use of a config file in SSIS?](https://mindmajix.com/ssis-interview-questions#config-file)

Config file in SSIS is used to provide inputs to connection managers different properties that packages use to assign values at runtime dynamically.

1. [What kind of containers can you use with SSIS packages?](https://mindmajix.com/ssis-interview-questions#containers) sequence, for loops, and for each loop.
2. [How to back up or retrieve the SSIS packages?](https://mindmajix.com/ssis-interview-questions#retrieve)

If your package is deployed on SQL Server then you can back up the MSDB database as all the packages on SQL server deploy at MSDB.

* 1. Can you use Lookup Transformation without using OLEDB Connection?

NA

2. What is the difference between Lookup and Merge Join Transformation?

Merge transformation is used to combine two sorted datasets (such as UNION operation).

Merge Join needs both inputs to be sorted.

SSIS Merge Join transformation is used to combine columns between different rows (such as SQL Joins(Full, Left, and inner join.)).

Look Up is used to find value(look for a value) in destination table with respect to source column values, i.e. check existence of source column value against destination column values.

Fuzzy Lookup transformation uses fuzzy matching to return one or more close matches in the reference table.

Fuzzy Grouping performs a data cleaning task by identifying rows of data that are likely to be duplicates.

Merge Join allows you to join to multiple columns based on one or more criterion, whereas a Lookup is more limited in that it only fetches a one or more values based on some matching column information -- the lookup query is going to be run for each value in your data source (though SSIS will cache the data source if it can).

Union All transformation can accept more than two datasets from input and combine.

3. How many outputs can Aggregate Transformation have?

The Aggregate transformation has one input and **one or more** outputs.

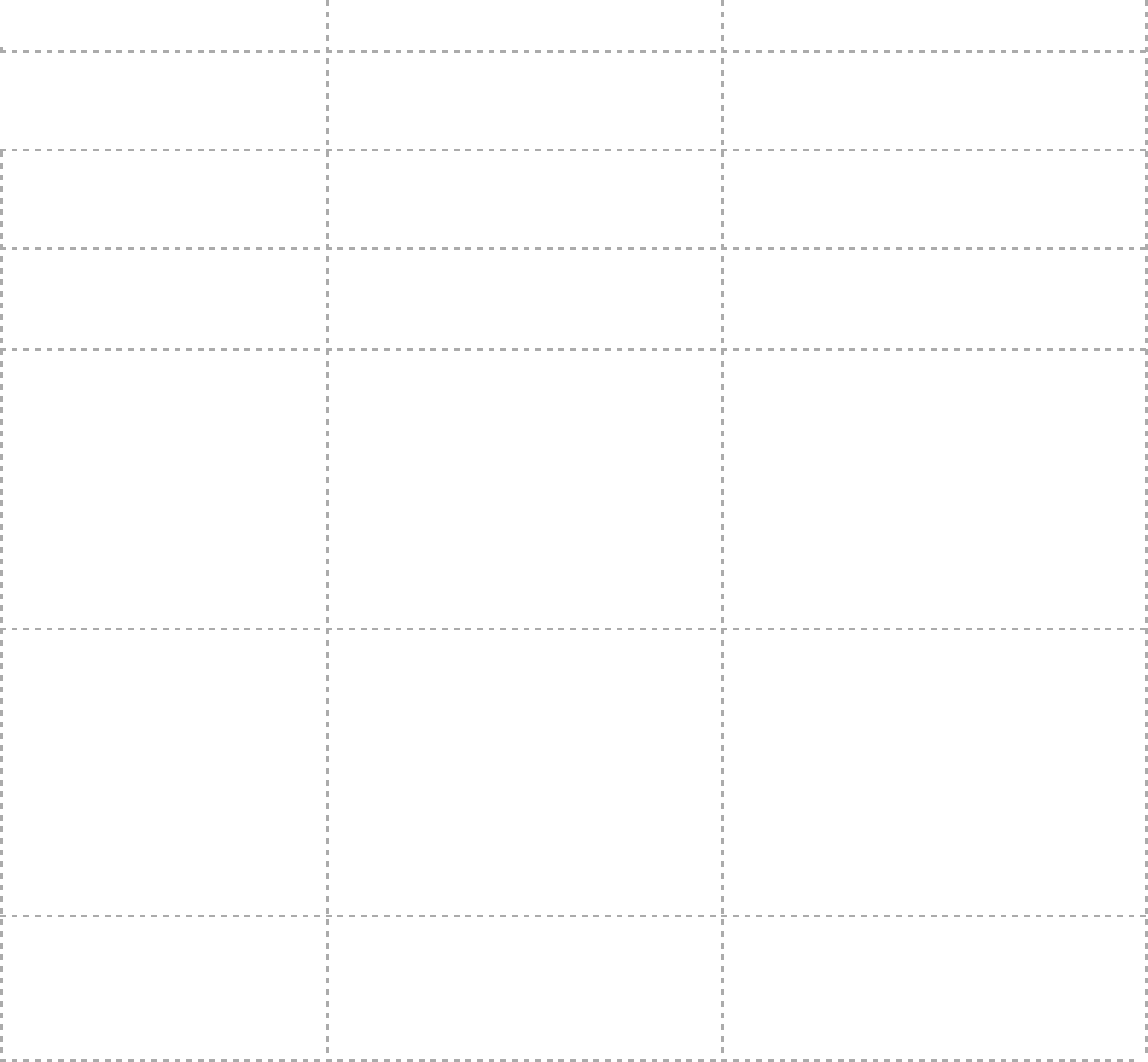
Be prepared with These questions and topic:

4. Differentiate between SSRS 2005 and SSRS 2008 ?

Differences between SSIS 2005 and 2008



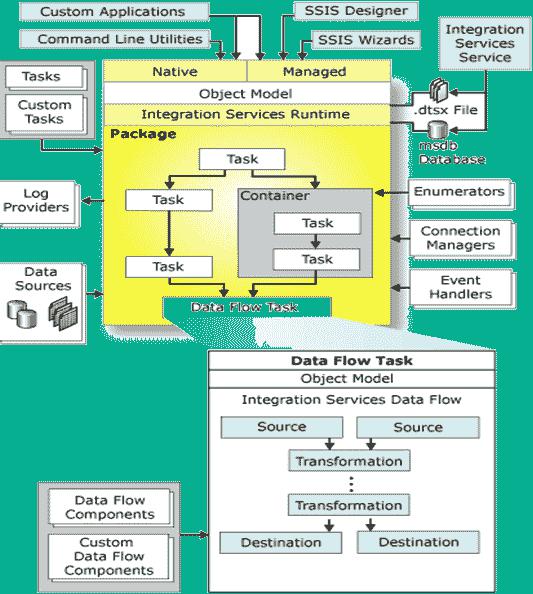
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SSIS 2005 |  | SSIS 2008 |  |  |
| **Script Task** | Here users can write the | | Here users can write the scripts | | |
|  | scripts in VB only. | | in C# and VB. | | |
| **DataProfilingTask** | DataProfilingTask is not in SSIS | | DataProfilingTask was introduced | | |
|  | 2005. |  | in SSIS 2008. |  |  |
| **Cache Transformation** | Cache Transformation is not in | | Cache Transformation was | | |
|  | SSIS 2005. |  | introduced in SSIS 2008. | | |
| **Lookup Transformation** | In SSIS-2005 for Error Output | | SSIS -2008 added an additional | | |
|  | look-ups had only the following feature “No match Out-Put” to | | | | |
|  | 3 options. |  | the SSIS 2005. | | |
|  | 1. | Fail Component |  |  |  |
|  | 2. | Ignore Failure |  |  |  |
|  | 3. | Redirect row |  |  |  |
| **Cache Mode** | Cache Mode is not in SSIS | | Cache Mode is introduced in SSIS | | |
|  | 2OO5. |  | 2OO8. 3-Different Cache Mode in | | |
|  |  |  | SSIS 2008: |  |  |
|  |  |  | 1. | FULL CACHE MODE | |
|  |  |  | 2. | PARTIAL CACHE | |
|  |  |  |  | MODE | |
|  |  |  | 3. | NO CACHE MODE. | |
| **OLEDB Provider for** | SSIS 2005 does not contains an SSIS 2008 contains an OLEDB | | | | |
| **Analysis Services** | OLEDB Provider for Analysis | | Provider for Analysis Services | | |
|  | Services |  |  |  |  |
|  |  |  |  |  |  |



Differences between SSIS 2008 and 2012

|  |  |  |
| --- | --- | --- |
| **Undo And Redo** | No Undo And Redo feature in | Undo And Redo feature |
|  | SSIS 2008 | available in SSIS 2012. |
| **SSIS PARAMETERS** | SSIS Parameters at package | SSIS Parameters at the package |
|  | level | level, task level and project |
|  |  | level. |
| **DQS TRANSFORMATION** No DQS in SSIS 2008. | | DQS Transformation is available |
|  |  | in SSIS 2012. |
| **Change Data Capture** | Introduced in SSIS 2008. But | CDC ControlTask available to |
| **(CDC)** | there is no task to support CDC | support CDC in SSIS 2012. |
|  | in SSIS 2008. |  |
| **Data Flow Tap** | No Data Tap Available in SSIS | Data Tap Available in SSIS 2012. |
|  | 2008. |  |

5. Architecture of SSIS?



6. How to organize a SSIS Package on production?

manifest file is the utility that can be used to deploy the package using the wizard on the file system and SQL Server database.

SSIS can be organized through

1. **Manifest we can organize it:**

a. Create deployment utility by setting its properties as true .

b. It will be formed in the bin folder of the solution once the package is built.

c. Copy all the files in the utility and use manifest file to display it on the Prod.

1. **Using DtsExec.exe utility.**

3)**Import Package directly in MSDB from SSMS by logging in Integration Services.**

7. Which programming language can be used to code embedded functions in SSRS?

**Visual Basic**

8. How to create Drill-Down reports?

Apply the Row Grouping from the parent Group to the child group.

9. How we can handle Early Arriving Facts.

Late-arriving dimensions (sometimes called early-arriving facts) occur when you have dimension data arriving in the data warehouse later than the fact data that references that dimension.

The fact row will be held in the suspense table until the associated dimension record has been processed.

1. What are the Command Line Utilities available in Reporting Service

Rsconfig(RS.EXE) utility: - script host that you can use to perform scripted operations.

Powershell cmdlets: - to configure and manage a report server connection to the report server database.

Rskeymgmt Utility– Encryption key management tool. You can use it to back up, apply, recreate, and delete symmetric keys. You can also use this tool to attach a report server instance to a shared report server database.

**Error handling: -**

In SSIS errors may occur because of the following reasons.

1. Data conversion errors
2. Lookup errors
3. Expression evaluation errors There are two types of errors in SSIS.
4. **Truncation**

A truncation generates results that might be usable. For example, if you try to insert 10 character values into 5 charter length columns.

1. **Error:**

This is a failure of the SSIS component and it generates the NULL result. For example: If you are trying to convert an alpha-numeric string to an integer. This will result in a data conversion error.

We can configure the SSIS component to handle the error and truncation in SSIS transformations.

**Fail Component:** The Data Flow task fails when an error or a truncation occurs. Failure is the default option for an error and a truncation.

**Ignore Failure:** The error or the truncation is ignored and the data row is directed to the output of the transformation or source.

**Redirect Row:** The error or the truncation data row is directed to the error output of the source, transformation, or destination.

**Event handlers: -**

Event Handlers can perform multiple tasks like:

* Cleaning up temp data storage or managing auditing history based on the package running or finishing an action.
* Retrieving and capturing system information before a package executes.
* Setting an email alert when a warning occurs or when a task fails.

OnError,OnInformation,OnPostExecute,OnPostValidate,OnProgress,OnTaskFailed,OnVariableValueChange d & OnWarning.

**SCD: -**

**This component is used if you want insert or update data records in dimension tables which changes slowly..(New, Changed or UnChanged)**

**If no record matches then it will treat as New Record or If record matches then it compares attributes for changed attributes if data looks updated then it updates record or if not then it leaves as it is unchanged.**

SCD is defined as slowly changing dimensions, and it applies to the cases where record changes over time.

There are three types of SCD and they are as follows:

SCD 1(Changing Attribute) – The new record replaces the original record used when there is no need to store historical data in the dimension table. SCD 2(Historical Attribute) – A new record is added to the existing customer dimension table

* Versioning
* Flagging
* Effective Date

SCD 3 – A original data is modified to include new data

only the current status and previous status of the row is maintained in the table.

To track these changes two separate columns are created in the table.

Type 4: Combination of types 1 and 2.

The current data is maintained in the dimension table and the history is stored in another table.

This type of dimension adds a lot of complexity. <https://www.folkstalk.com/2012/03/slowly-changing-dimensions-scd-types.html>

**SCD Type Summary**

**Type 0** **Ignore any changes and audit the changes.**

**Type 1** **Overwrite the changes**

**Type 2** **History will be added as a new row.**

**Type 3** **History will be added as a new column.**

**Type 4** **A new dimension will be added**

**Type 6** **Combination of Type 2 and Type 3**

[**https://www.sqlshack.com/implementing-slowly-changing-dimension**](https://www.sqlshack.com/implementing-slowly-changing-dimensions-scds-in-data-warehouses/)[**s-scds-in-data-warehouses/**](https://www.sqlshack.com/implementing-slowly-changing-dimensions-scds-in-data-warehouses/)

**Type 2 and Type 6 are the most commonly used dimension in a data warehouse.**

**logging mode property.**

**Disabled,Enabled,Use Parent Setting**

**Data viewers**

**Column Chart**

**Grid**

**Histogram**

**Scatter Plot**

**Package level and project level deployment: -**

Package Level Deployment(SQL Server 2005/8): - It will store the packages to the file system or msdb database. Package configurations files need to be done at the package level. Project parameters are not supported in this.

Convert to Package Deployment Model.

CreateDeploymentUtility property to True

AllowConfigurationChanges property is a key setting as well, and when set to True,

project level deployment(SQL 2012 onwards): - Right click on the project and deploy it will deploy all the packages in single shot. It will create an ispac file and deploy it to the server. It will store the ssis catalog.

How to achieve parallelism in SSIS?

Parallelism is achieved using the MaxConcurrentExecutable property of the package. Its default is -1 and is calculated as the number of processors + 2.

Using the MaxConcurrent Executable property of the package. Its default is -1 and it is calculated as a number of processors +2

**What is the Data Path and how is it different from a Precedence Constraint? Answer:** Data Path is used in a Data Flow task to connect to different components of a Data Flow and show the transition of the data from one component to another. A data path contains the Meta information of the data fowing through it, such as the columns, data type, size, etc. When we talk about differences between the data path and precedence constraint; the data path is used in the data fow, which shows the fow of data. Whereas the precedence constraint is used in control fow, which shows control fow or transition from one task to another task

What is the RetainSameConnection property and what is the impact?

Whenever a task uses a connection manager to connect to source or destination database, a connection is opened and closed with the execution of that task. Sometimes you might need to open a connection, execute multiple tasks and close it at the end of the execution. This is where the RetainSameConnection property of the connection manager might help you. When you set this

property to TRUE, the connection will be opened on the first time it is used and remain open until execution of the package completes.

**How to create a temp table in SSIS?**

A temporary table can be created in SSIS through the Execute SQL task. For this, the **RetainSameConnection** has to be set to True so that the table can be created in the **Control Flow task**. This table can be retained in another task.

What is the use of Bulk Insert Task in SSIS?

Bulk Insert Task is used to upload large amounts of data from flat files into Sql Server. It supports only OLE DB connections for destination databases.

6. difference between OLEDB destination and SQL Server destination.

ans : OLEDB destination loads the records in batches, whereas SQL Server destination loads all the records at one go..

1. Have you used sort transformation? Why did you use that? (Or) why have you not used that?
   1. How many types of transformations? What are they?
   2. The difference between merge join and lookup? which is better? why not merge join.
2. In order to speed up the loading of millions of records from the source file what should i do.? We have few the below properties in SSIS

Avoid Sort T/F wherever possible

Drop all Non-Clustered Indexes and Clustered Index on the destination table Avoid SELECT \*--more the columns less no of rows in the buffer

OLEDB Destination Settings: - fast load,Keep Identity ,Table Lock,Check Constraints(recommendation is to un-check it if you are sure that the incoming data is not going to violate constraints of the destination table.),Rows per batch ,Maximum insert commit size,DefaultBufferSize(small as 64 KB and as large as 100 MB),DefaultBufferMaxRows

Keep Identity : -

By default this setting is unchecked which means the destination table (if it has an identity column) will create identity values on its own.

Keep Nulls – Again by default this setting is unchecked which means default value will be inserted (if the default constraint is defined on the target column) during insert into the destination table if NULL value is coming from the source for that particular column. If you check this option then the default constraint on the destination table's column will be ignored and the preserved NULL of the source column will be inserted into the destination.

Rows per batch:(default value for this setting is -1)

Each batch is copied to the transaction log, which will cause it to grow quickly, but offers the ability to back up that transaction log after each batch. Also, having a large batch will negatively affect memory if you have indexes on the target table, and if you are not using table locking, you might have more blocking going on.

Maximum insert commit size

Changing the default value for this setting will put overhead on the dataflow engine to commit several times. but at the same time it will release the pressure on the transaction log and tempdb to grow specifically during high volume data transfers.

<http://msbijobs.blogspot.com/p/company-wise-interview-questions.html>

#7 - DefaultBufferSize and DefaultBufferMaxRows :

The execution tree creates buffers for storing incoming rows and performing transformations.

The number of buffers created is dependent on how many rows fit into a buffer and how many rows fit into a buffer dependent on a few other factors. The first consideration is the estimated row size, which is the sum of the maximum sizes of all the columns from the incoming records. The second consideration is the DefaultBufferMaxSize property of the data flow task. This property specifies the default maximum size of a buffer. The default value is 10 MB and its upper and lower boundaries are constrained by two internal properties of SSIS which are MaxBufferSize (100MB) and MinBufferSize (64 KB). It means the size of a buffer can be as small as 64 KB and as large as 100 MB. The third factor is, DefaultBufferMaxRows which is again a property of data flow task which specifies the default number of rows in a buffer. Its default value is 10000.

If the size exceeds the DefaultBufferMaxSize then it reduces the rows in the buffer. For better buffer performance you can do two things.

First you can remove unwanted columns from the source and set data type in each column appropriately, especially if your source is a flat file. This will enable you to accommodate as many rows as possible in the buffer.

Second, if your system has sufficient memory available, you can tune these properties to have a small number of large buffers, which could improve performance. Beware if you change the values of these properties to a point where page spooling (see Best Practices #8) begins, it adversely impacts performance. So before you set a value for these properties, first thoroughly test in your environment and set the values appropriately.

1. Better performance with parallel execution
2. When to use event logging and when to avoid it.
3. Monitoring the SSIS Performance with Performance Counters Launch Performance Monitor:
4. Start -> All Programs -> Administrative Tools -> Performance
5. Load the SSIS related Counters

In the Performance Object, select SQL Server:SSIS Pipeline and SQL Server:SSIS Service.

SSIS provides a set of performance counters. Among them, the following few are helpful when you

tune or debug your package:

Buffers in use

Flat buffers in use

Private buffers in use

Buffers spooled

Rows read

Rows written

“Buffers in use”, “Flat buffers in use” and “Private buffers in use” are useful to discover leaks. During package execution time, we will see these counters fluctuating. But once the package finishes execution, their values should return to the same value as what they were before the execution. Otherwise, buffers are leaked.

“Buffers spooled” has an initial value of 0. When it goes above 0, it indicates that the engine has started memory swapping. In a case like this, set Data Flow Task properties BLOBTempStoragePath and BufferTempStoragePath appropriately for maximal I/O bandwidth.Buffers Spooled: The number of buffers currently written to the disk. If the data flow engine runs low on physical memory, buffers not currently used are written to disk and then reloaded when needed.

“Rows read” and “Rows written” show how many rows the entire Data Flow has processed.

12. FastParse property

The Fast Parse option in SSIS can be used for very fast loading of flat file data. It will speed up parsing of integer, date and time types if the conversion does not have to be locale-sensitive. This option is set on a per-column basis using the Advanced Editor for the flat file source.

**Define data warehouse: -**

Data warehousing (DW) is a method of gathering and [analyzing data](https://www.interviewbit.com/data-analyst-interview-questions/) from many sources in order to get useful business insights

**Star Schema,Snowflake Schema**

1. [What is the difference between Char and VARCHAR](http://www.techbrothersit.com/2016/02/what-is-difference-between-char-and.html)
2. [What is difference between VARCHAR and NVARCHAR](http://www.techbrothersit.com/2016/02/what-is-difference-between-varchar-and.html)
3. [What are different Types of Constraints Available in SQL Server](http://www.techbrothersit.com/2016/02/what-are-different-types-of-constraints.html)
4. [How to Enable Change Data Capture(CDC) on Database in SQL Server](http://www.techbrothersit.com/2013/06/change-data-capture-cdc-sql-server.html)
5. [How to Disable Change Data Capture(CDC) on Database in SQL Server](http://www.techbrothersit.com/2013/06/change-data-capture-cdc-sql-server_6.html)

**What is Data Mining?**

Data mining is analyzing data from different perspectives, dimensions, and patterns and summarizing them into meaningful content. Data is often retrieved or queried from the database in its format. On the other hand, it can be defined as the method or process of turning raw data into useful information.

**Data Lake**

A data lake is a centralized repository for structure and unstructured data storage. It can be used to store raw data without any structure schema, and there is no need to perform any ETL or transformation job. Any type of data can be stored here, like images, text, files, and videos, and even it can store machine learning model artifacts, real-time and analytics output, etc. Data retrieval processing can be done via export, so the schema is defined on reading. It mainly stores raw and unprocessed data. The main focus is to capture and store as much data as possible.

**What is Data Modelling?**

Data Modelling is a very simple step of simplifying an entity here in the concept of data engineering. It will simplify complex software by simply breaking it up into diagrams and further breaking it into flow charts. Flowcharts are a simple representation of how a complex entity can be broken down into a simple diagram. This will give a visual representation, an easier understanding of the complex problem, and even better readability to a person who might not be proficient in that particular software usage.

**What is a dimensional model in a data warehouse?**

A dimensional model is a design approach for organizing data in a data warehouse. It consists of fact tables and dimension tables. Fact tables store quantitative data (e.g., sales, cost, revenue) and are typically linked to one or more dimension tables, which store descriptive data (e.g., product, customer, time). Dimensional modeling lets users quickly understand and analyze data by breaking it down into smaller, more manageable pieces.

**What is a Fact? What are the types of Facts?**

A fact may be a central component of a multi-dimensional model that contains the measures to be analyzed. Facts are related to dimensions.

Types of facts are:

* Additive Facts
* Semi-additive Facts
* Non-additive Facts

**What is the Data Pipeline?**

Data Pipeline refers to any set of process elements that move data from one system to a different one. Data Pipeline is often built for an application that uses data to bring value. It is often used to integrate the info across the applications, build info-driven web products, and complete data mining activities. Data engineers build the data pipeline.

**What is Dimensional Data Modeling?**

Dimensional modeling is a set of guidelines to design database table structures for easier and faster data retrieval. It is a widely accepted technique. The benefits of using dimensional modeling are its simplicity and faster query performance. Dimension modeling elaborates logical and physical data models to further detail model data and data-related requirements. Dimensional models map the aspects of every process within the business.

Dimensional Modelling is a core design concept used by many data warehouse designers design data warehouses. During this design model, all the info is stored in two sorts of tables.

* Facts table
* Dimension table

The fact table contains the facts or measurements of the business, and the dimension table contains the context of measurements by which the facts are calculated. Dimension modeling is a method of designing a data warehouse.

Types of Dimensional Modelling are listed below:

* Conceptual Modelling
* Logical Modelling
* Physical Modelling

scd

lokup(how to select full cachemode and partial cache mode)

fuzzy lookup

properties in ssis package

sql server versions

2 tables how to find the records not in the first or second table.

Environmental variables

isolation levels

parameters in ssis catalog

statements inside the triggers

have you used delay validation

parallel load in ssis and how to enable it

how to handle performance issues in sql, ssis

what is the difference between lookup & merge transformation? have you handle any performance issue in ssis package functions used in derived column transformation

Explain the types of Perfmon counters for SSIS? Differentiate Execute TSQL Task and Execute SQL Task. What do you mean by Process Bytes in SSIS?

Process bytes represent the amount of memory presently used by integration services.

what is, dimensional modeling

types of fact tables.

coding standards in sql, ssis.

reverse the string in sql

data modeling and data mining

what id mdx

diff between tabular and multidimensional modelsntional

**—-------------------------------------HR—-----------------**

In HR interviews these questions are frequently asked.

1. Tell me something about yourself? Or Introduce yourself.

MY Name is….

*I am from…. Currently I am staying in* ….

*I am a graduate in…*

* *am quite good at…*. *Thank you.*
  1. Why should we hire you for this position?
* *believe I have the required skills and experience that are required for the position. Based on the skills that I have learned in the current company and the automations solutions I have implemented on my own, I feel that I perfectly fit this role.*
  1. What are your strengths?
* am the focused person who always works with positive attitude to achieve the objectives of the company

Honesty, team player, confidence, dedication etc

* 1. Tell me about your weaknesses.
* won't feel comfortable if I haven't finished my work.

Leaving things/tasks halfway is difficult for me Trusting people easily, I never want to hurt people.

helping nature and being good with everyone.

5. Why do you want to join our company?

Accenture is a multinational company that offers professional services in the field of

Information technology and consulting.

A unique, diverse workplace environment

upskilling their employees(Continuous learning and development opportunities) Career Counselor.

6. What do you look for in a job?

How much I am learning compared to yesterday?

What change I have made in the process I am following compared to yesterday.

Passion

Satisfaction

Work life balance

1. What are your career goals?
2. Where do you see yourself in 5 years from now?

I would be more experienced not in skill set but in mindset. Because, i believe literally anybody can learn the skills through online or books. But, professional experiences, new ideologies, mature mindset, proper behavior, work under pressure, handling deadlines, dealing with complex situations, there are plenty of things I would be more experienced in by meeting people, listening, observing them, and finally make me mentally strong enough to stand amid the adversities.”

9. What type of salary do you expect for this job?

*market standards*

*Money is important to me, but is not my main concern.*

*Opportunity and growth are far more important. What I’d rather do, if you don’t mind, is explore if I’m right for the position, and then talk about money. Would that be okay?*

*I’m sure the company has already established a salary range for this position. Could you tell me what that is?*

1. Give me an example of your Initiative. Can give on etl monitoring automation and mail alerts.
2. Why did you choose a career in this field?
3. What if I say that you are rejected?

Being a experience candidate, I do understand that I might lack in certain areas which is expected for this position, I would go back and make sure, what needs to be done and learn required things and hopefully attend an interview again with you guys after 6 months

13. Describe an ideal working environment for you in an organization.

all the team members feel a sense of ownership of the work being done, and work together

towards an optimistic, realistic goal.

“peace of mind”,“Satisfaction” .

14. What kind of people do you find difficult to work with?

with insufficient experience

getting something done fast over getting it done right don’t listen and they don’t really care what you think

1. Have you ever failed in any of your decisions? How and why?
2. [**Why do you want to leave your current company and join us"?**](https://www.quora.com/What-is-the-best-answer-for-Why-do-you-want-to-leave-your-current-company-and-join-us?no_redirect=1)
   1. I am looking for a bigger place where my current level of skills can be utilized to full potential
   2. I am looking for assuming more responsibility and leadership and I see bright chances for the same in your company
   3. My last company has helped me enormously in honing my skills and my contribution has been consistent with what I learnt there. Now, I think, I am ready for the next stage of my career and your company offers a matching platform

Here are some common behavioral interview questions you might encounter:

* Tell me a time when you made a mistake.
* Tell me a time when you disagreed with your co-worker or boss.
* Tell me a time when you played a critical role on a team.
* Tell me a time when you had to communicate a difficult topic.
* Tell me about a time that you generated change in your organization.
* What is the most difficult situation you have ever faced in your life? How did you handle it?
* Describe a process that you made better using a simple solution.

* Give me an example of when you had to use your negotiating skills.
* Describe a time where you went above and beyond your job description.
* Tell me about a time when you had to think of more than one way to convince a person to agree with an idea.
* Tell me about a time when you had a conflict with a person and how you handled it.
* Tell me about a time when you had a group conflict. How did you overcome this conflict?
* Tell me a time when you motivated employees.