

SQL SERVER INTEGRATION SERVICE CHEAT SHEET

SSIS BASICS

SSIS

SSIS is a component of Microsoft SQL Server database which can be used to perform data migration tasks, by collecting the data from various data sources and storing it in the central location. The main function performed by SSIS is Extraction, Transformation and Loading

- **Extraction:** It is the collection of data from various sources
- **Transformation:** It is a different form of data collected from different sources and is converted according to the Business requirement
- **Loading:** It is a different form of data collected from different sources and is converted according to the Business requirement

Features of SSIS

- **Studio environment:** SSIS includes two studios
 - o **SSDT:** Referred to as SQL Server Data Tools which is used to develop integration service packages and perform several
 - o **SSMS:** Referred to as SQL Server management studio which is used for the package management in a production environment and perform tasks
- **Packages:** SSIS package is a collection of control flow and data flow
- **Expressions:** SSIS Expressions are the combination of literals, identifiers and operators
- **Event Handling:** Handling a workflow that is designed for the number of events in different ways where it might occur. It is similar to building the packages

Data Warehouse

It captures the data from various sources for useful analysis and access

Data Warehousing

It is a collection of data accumulated which is used for assembling and managing data from various sources for better decision making in the Business

Transformations

Aggregate - It is available in SSIS is used to apply aggregate functions to Record Sets to produce new output records from aggregated values.

Audit - It is used to add Package and Task Level

Metadata - such as Machine Name, Execution Instance, Package Name, Package ID, etc.

Character map - It is used to Performs SQL Server operations such as changing data from lower case to upper case.

Conditional split - It is used to separate available input into separate output pipelines based on Boolean Expressions designed for each output.

Copy column - It is used to add a copy of column to the output so that we can later transform the copy keeping the original for auditing

Cache: It is used to write the data from connected data source to a cache file

Data conversion - It is used to convert columns data types from one type to another type. It stands for Explicit Column Conversion.

Data mining query - It is used to perform data mining query against analysis services and to manage Predictions, Graphs and Controls.

Derived column - Used to create a new column from given expressions.

Export column - It is used to export an Image specific column from the database to a flat file.

Fuzzy grouping - It is used for data cleansing by finding rows that are likely duplicates.

Fuzzy lookup - It is used for Pattern Matching and Ranking based on fuzzy logic.

Import column - It reads image specific column from database onto a flat file.

Lookup - It performs search of a given reference object set against a data source and is used for exact matches only.

Multi cast - It sends a copy of supplied Data Source onto multiple Destinations.

Merge - It is used to Merges two sorted data sets into a single data set into a single data flow.

Merge join - It is used to merge two data sets into a single dataset using a join junction.

Row count - It is used to store the resulting row count from the data flow / transformation into a variable.

Row sampling - It captures the sample data by using a row count of the total rows in dataflow specified by rows or percentage.

Union all - It is used to Merge multiple data sets into a single dataset.

Pivot - It is used for data source normalization by converting rows into columns

Unpivot - It is used for de-normalizing the data structure by converting columns into rows in case of building Data Warehouses.

Uses of SSIS

- It is used to combine data from multiple data sources
- Populating data warehouses and data marts
- For Cleaning and standardizing data
- To Build BI into the data transformation process
- Used for automating administrative functions and loading the data

Statement	Expression
Create a file name with today's date	Expression on the Flat File or File Connection Manager: "C:\Project\MyExtract"+(DT_WSTR,30)(DT_DBDATE)GETDATE() + ".csv" Expression Output Example: C:\Project\MyExtract 2009-03-20.csv
Use a 2 digit date	RIGHT("0" + (DT_WSTR, 2)MONTH(GETDATE()),2) Expression Output: 03 (if the month is March)
Multiple condition if statement	In this example, the statement determines that if the ColumnName is blank or NULL, it will be set to unknown. To make a Logical AND condition, use "&&" instead of the " " operator. ISNULL(ColumnName) TRIM(ColumnName)="" ? "Unknown" : ColumnName
Returns the first five characters from a zip code	Derived Column Transform in the Data Flow: SUBSTRING(ZipCodePlus4,1,5)
Remove a given character from a string	Derived Column Transform in the Data Flow: REPLACE(SocialSecurityNumber, "-", "")
Uppercase data	Derived Column Transform in the Data Flow: UPPER(ColumnName)
Replace NULL with another value	Derived Column Transform in the Data flow: ISNULL(ColumnName)?"New Value": ColumnName
Replace blanks with NULL values	Derived Column Transform in the Data Flow: TRIM(ColumnName)="" ? (DT_STR, 4, 1252) NULL(DT_STR, 4, 1252) : ColumnName

Remove any non-numeric data from a column

```
Script Transform in the Data Flow Task with the code as follows:
Imports System.Text.RegularExpressions
Public Overrides Sub Input()_ProcessInputRows
    (ByVal Row As Input)Buffer
    If Row.ColumnName_IsNull = False Or Row.
    ColumnName="" Then
        Dim pattern As String = String.Empty
        Dim r As Regex = Nothing
        pattern = "[^0-9]"
        r = New Regex(pattern, RegexOptions.Compiled)
        Row.ColumnName =
        Regex.Replace(Row.ColumnName, pattern, "")
    End If
End Sub
```

Convert text to proper case

Script Transform with the line of partial code as follows: Row.OutputName = StrConv(Row.Input Name, VBStrConv.ProperCase)

Build dynamic SQL statement

Expression on the SQLStatementSource property of Execute SQL Task(VB 2008):
"SELECT Column From " + @[User::TableName] +
WHERE DateFilterColumn = "" + (DT_WSTR,4)YEAR(@
[User::DateTimeVar]) + RIGHT("0" + (DT_WSTR,2)
MONTH(@
[User::DateTimeVar]),2) + RIGHT("0" + (DT_WSTR,2)
DAY(@
[User::DateTimeVar]),2) + ""

Calculate beginning of the previous month

Expression on Derived Column Transform:
ROUND(YourNumber, 2)
Expression Output Example: 1.260000

Round to the nearest two decimal mark

Expression on Derived Column Transform:
ROUND(YourNumber, 2)
Expression Output Example: 1.260000

