Integration Services

Mag. Thomas Griesmayer

- It is a process used in data warehousing and data integration to move data from one or more sources into a target.
- Extract: This step involves retrieving data from various source systems. These sources can be databases, flat files, web services, or any other data storage. The goal is to gather data from different systems into a single repository.
- Transform: During the transformation step, the extracted data is cleaned, enriched, and transformed into a format that is suitable for analysis or further processing. This can include filtering, aggregating, joining, or splitting data, as well as applying business rules and data validations.
- Load: In the load step, the transformed data is loaded into the target system, such as a database, data warehouse, or data mart. This step may involve inserting, updating, or deleting data in the target system to ensure it reflects the latest information.

- ETL processes are crucial for integrating data from multiple sources, maintaining data quality, and enabling business intelligence and analytics applications. They ensure that the data is accurate, consistent, and readily available for analysis.

Integration Services

- SQL Server Integration Services (SSIS), is a platform for building enterprise-level data integration and data transformation solutions (ETL).
- Data Integration: integration of data from various sources.
- Data Transformation: set of transformation capabilities, allowing users to clean, aggregate, merge, and reshape data.
- Data Loading: it can load it into a variety of destinations.
- Workflow and Task Automation: SSIS includes a workflow engine that can orchestrate various tasks, such as executing SQL statements, sending email notifications, or running scripts.
- Error Handling and Logging: It can capture and log errors, allowing for easier troubleshooting and ensuring data integrity throughout the ETL process.
- Scalability and Performance: SSIS is designed to handle large volumes of data efficiently includeing parallel execution of tasks, data partitioning, and efficient memory management.
- Graphical Development Environment: users can design, test, and deploy ETL packages.
- Extensibility: SSIS can be extended using custom scripts written in C# or VB.NET.

Execute SQL Task

- An Execute SQL Task in SQL Server Integration Services (SSIS) is a task that allows you to run SQL statements during the execution of an SSIS package.
- This task is highly versatile and can be used for various purposes, such as querying databases, manipulating data, creating or altering database objects, and controlling the flow of the package.



```
DROP TABLE IF EXISTS dbo.CUSTOMER_VALUE;

CREATE TABLE dbo.CUSTOMER_VALUE
(
    CUSTOMER_VALUE_ID INTEGER PRIMARY KEY,
    CUSTOMER_VALUE        VARCHAR(50),
    CUSTOMER_MARGIN       DECIMAL(10,2)
);
```

https://chatgpt.com (22.5.2024)

- Download Files: Transfer files from an FTP server to a local directory.
- Upload Files: Transfer files from a local directory to an FTP server.
- Delete Files: Remove files from an FTP server.
- Create Directories: Create new directories on an FTP server.
- Remove Directories: Delete directories on an FTP server.

https://chatgpt.com (22.5.2024)
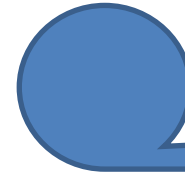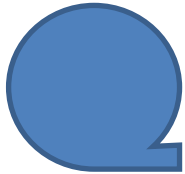
- The File System Task is used to perform operations on files and directories within the file system. This task is useful for automating common file-related operations, such as copying, moving, deleting, or renaming files and directories as part of an ETL process.
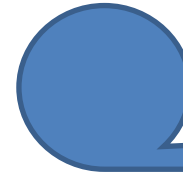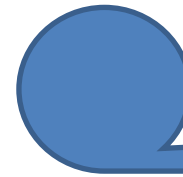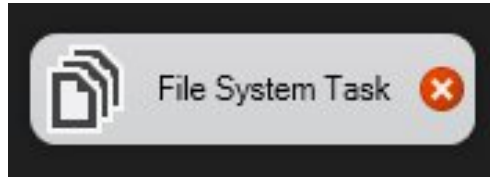
- The Data Flow Task is responsible for extracting, transforming, and loading (ETL) data from various sources to various destinations.
- During the execution of a Data Flow Task, data moves through the pipeline from the sources, through the transformations, and finally into the destinations.
- The process is highly efficient, allowing large volumes of data to be processed quickly.

https://chatgpt.com (22.5.2024)
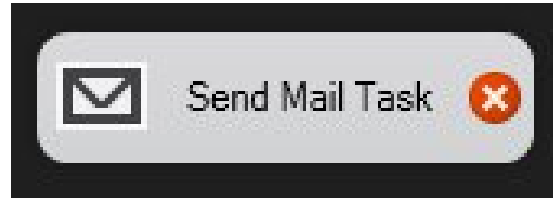
- It allows you to call a web service method and retrieve the results for use in your ETLProcess.
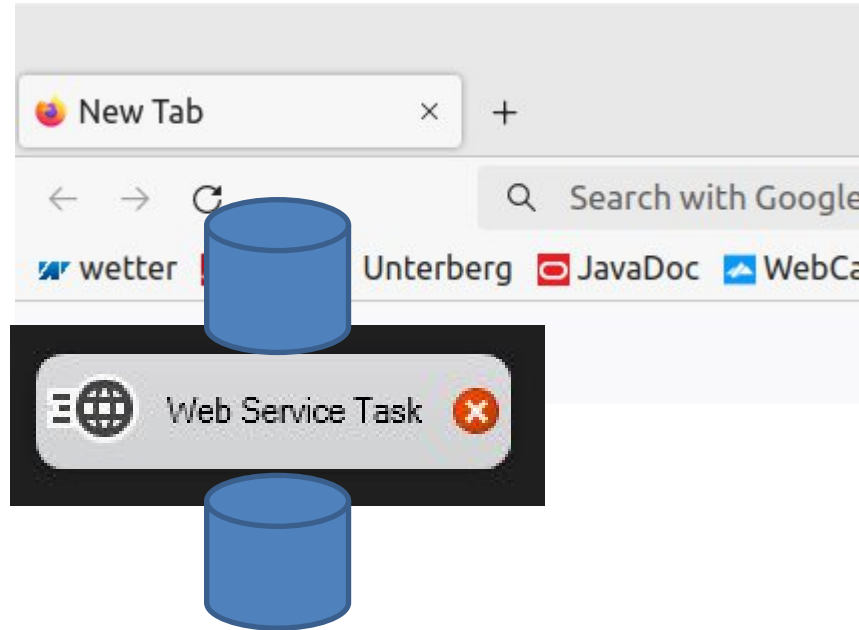
- This can be useful for notifications, alerts, or sharing results of data processing tasks.
- The task uses an SMTP (Simple Mail Transfer Protocol) server to send the emails.

- The Foreach Loop Container in SSIS is a versatile tool that enhances the automation and efficiency of ETL processes by allowing repeated execution of tasks over various types of collections.
- Its flexibility in handling different enumerators makes it suitable for a wide range of scenarios, from file processing to database operations.

Foreach Loop
Container

https://chatgpt.com (22.5.2024)
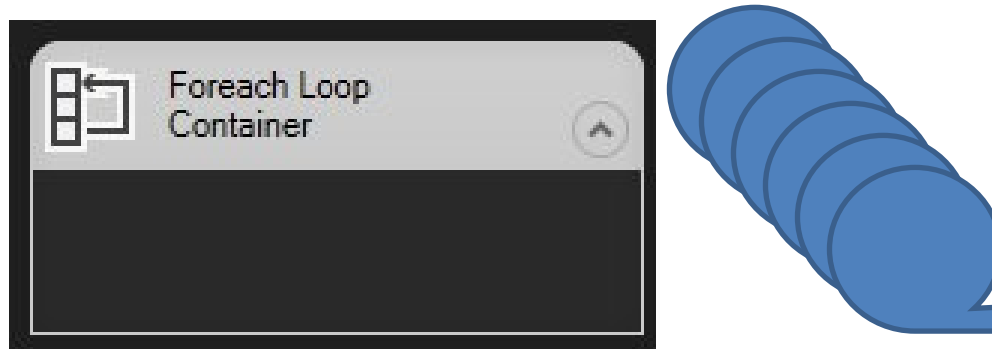
- In SQL Server Integration Services (SSIS), a Data Source is a component that provides a connection to various types of data stores for the purpose of extracting data. Data sources are integral to the ETL (Extract, Transform, Load) process, as they allow SSIS packages to access and retrieve data from diverse sources.

- Datasorces:
    - Relational Databases
    - Flat Files
    - Excel Files
    - XML Files
    - ADO.NET Sources
    - Other Data Sources
    - Web Services
    - SharePoint Lists

https://chatgpt.com (22.5.2024)

Datasource

Flat File Source   OLE DB Source ⊗   XML Source ⊗   Excel Source ⊗

1,Susi
2,Werner
3,Hans

```
CREATE TABLE
(
    CID INTEGER,
    NAME
VARCHAR(10)
)

1,Susi
2,Werner
3,Hans
```

```xml
-<CATALOG>
 -<CD>
    <TITLE>Empire Burlesque</TITLE>
    <ARTIST>Bob Dylan</ARTIST>
    <COUNTRY>USA</COUNTRY>
    <COMPANY>Columbia</COMPANY>
    <PRICE>10.90</PRICE>
    <YEAR>1985</YEAR>
  </CD>
 -<CD>
    <TITLE>Hide your heart</TITLE>
    <ARTIST>Bonnie Tyler</ARTIST>
    <COUNTRY>UK</COUNTRY>
    <COMPANY>CBS Records</COMPANY>
    <PRICE>9.90</PRICE>
    <YEAR>1988</YEAR>
  </CD>
```

https://chatgpt.com (22.5.2024)

- The pipe concept is a fundamental paradigm in data processing and integration, providing a robust framework for building modular, efficient, and scalable systems.
- In SSIS, this concept is realized through the Data Flow Task, where data flows seamlessly through sources, transformations, and destinations, embodying the principles of modularity, reusability, and flexibility.

Sort

CID	Name
4	Alex
1	Fritz
3	Clara
2	Barbara
5	Doris
6	Max
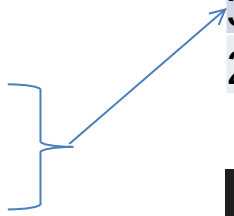
CID	Name
3	Clara
2	Barbara

Internal

CID	Name
2	Barbara
3	Clara

CID	Name
4	Alex
1	Fritz

↓↑ Sort ✖

CID	Name
4	Alex
2	Barbara
3	Clara
1	Fritz

CID	Name
4	Alex
2	Barbara
3	Clara
5	Doris
1	Fritz
6	Max

Merge Join

CUSTOMER_ID	Name	Gender
2	Susi	f
1	Fritz	m
3	Hans	m
4	Alex	m

Gender	Gender_Name
f	Female
M	Male

Merge Join
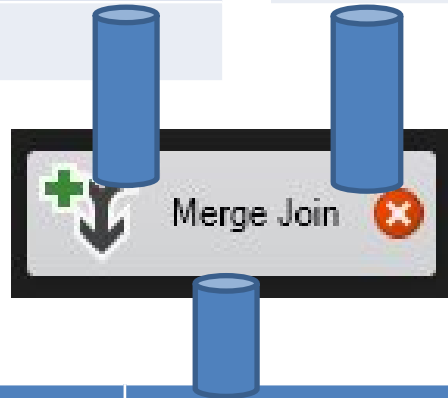
CUSTOMER_ID	Name	Gender	Gender_Name
2	Susi	f	Female
1	Fritz	m	Male
3	Hans	m	Male
4	Alex	m	Male

Union All

CUSTOMER_ID	Name	Gender
2	Susi	f
1	Fritz	m
3	Hans	m
4	Alex	m

CUSTOMER_ID	Name	Gender
5	Andi	f
6	Max	m

Union All

CUSTOMER_ID	Name	Gender
2	Susi	f
1	Fritz	m
3	Hans	m
4	Alex	m
5	Andi	f
6	Max	m

Derived Column

CUSTOMER_ID	Name	Gender	Birth_Date
2	Susi	f	19801210
1	Fritz	m	20000101
3	Hans	m	19991117
4	Alex	m	19801104
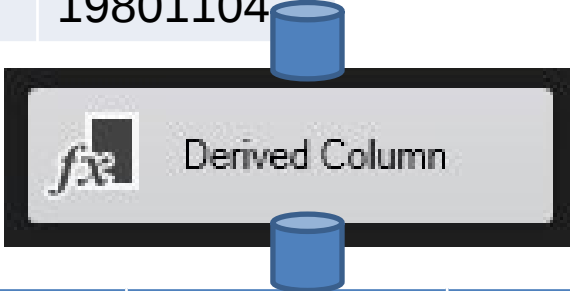
SUBSTRING(...)

fx Derived Column

CUSTOMER_ID	Name	Gender	Birth_Date	Birth_Date2
2	Susi	f	19801210	1980-12-10
1	Fritz	m	20000101	2000-01-01
3	Hans	m	19991117	1999-11-17
4	Alex	m	19801104	1980-11-04

Name	Gender	Birth_Date	Birth_Date2
Susi	f	19801210	1980-12-10
Fritz	m	20000101	2000-01-01
Hans	m	19991117	1999-11-17
Alex	m	19801104	1980-11-04

Data convertion

Data Conversion
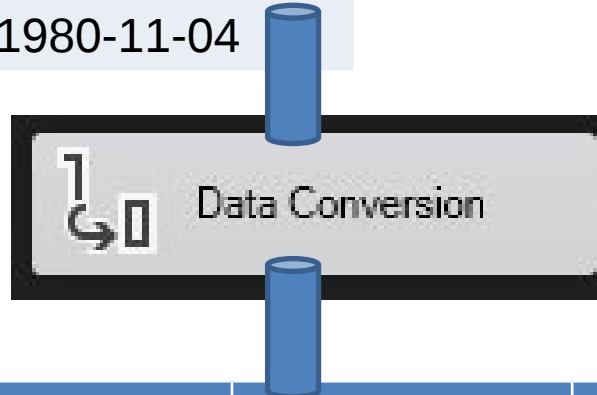
CAST DATE

CUSTOMER_ID	Name	Gender	Birth_Date	Birth_Date2
2	Susi	f	19801210	1980-12-10
1	Fritz	m	20000101	2000-01-01
3	Hans	m	19991117	1999-11-17
4	Alex	m	19801104	1980-11-04

Conditional Split

Name	Gender	Birth_Date	Birth_Date2
Susi	f	19801210	1980-12-10
Fritz	m	20000101	2000-01-01
Hans	m	19991117	1999-11-17
Alex	m	19801104	1980-11-04

Conditional Split
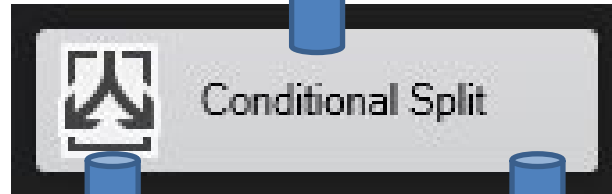
GENDER="m"

Name	Gender	Birth_Date	Birth_Date2
Susi	f	19801210	1980-12-10

Name	Gender	Birth_Date	Birth_Date2
Fritz	m	20000101	2000-01-01
Hans	m	19991117	1999-11-17
Alex	m	19801104	1980-11-04

Name	Gender	Birth_Date	Birth_Date2
Susi	f	19801210	1980-12-10
Fritz	m	20000101	2000-01-01
Hans	m	19991117	1999-11-17
Alex	m	1980110	1980-11-04

Aggregate

Aggregate
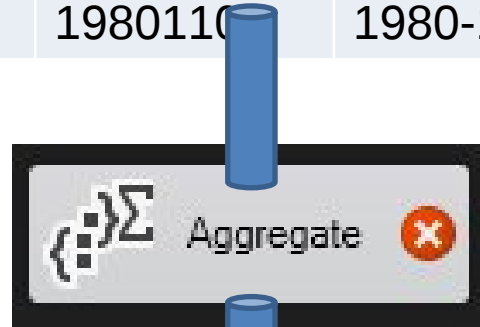
GROUP BY Gender
count(*)

GENDER	Name
m	3
F	1

Integration Services

CUSTOMER_ID	Name
5	Max
6	Alex

Recordset Destination ✕

Internal

CUSTOMER_ID	Name
2	Susi
1	Fritz
3	Hans
4	Alex

CUSTOMER_ID	Name
2	Susi
1	Fritz
3	Hans
4	Alex
5	Max
6	Alex

Data Destination

Name	Gender	Birth_Date	Birth_Date2
Susi	f	19801210	1980-12-10
Fritz	m	20000101	2000-01-01
Hans	m	19991117	1999-11-17
Alex	m	19801104	1980-11-04

Flat File Destination ❌

OLE DB Destination ❌

Excel Destination ❌