

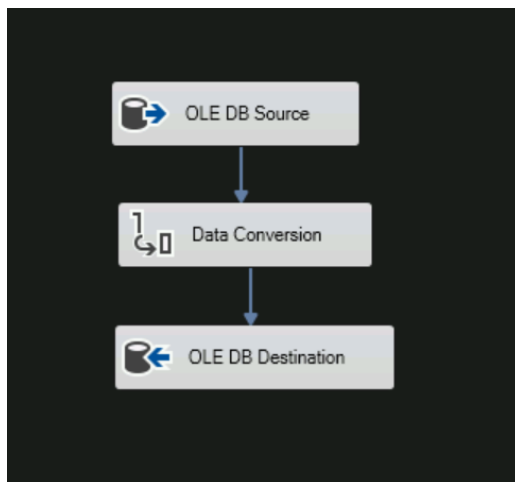
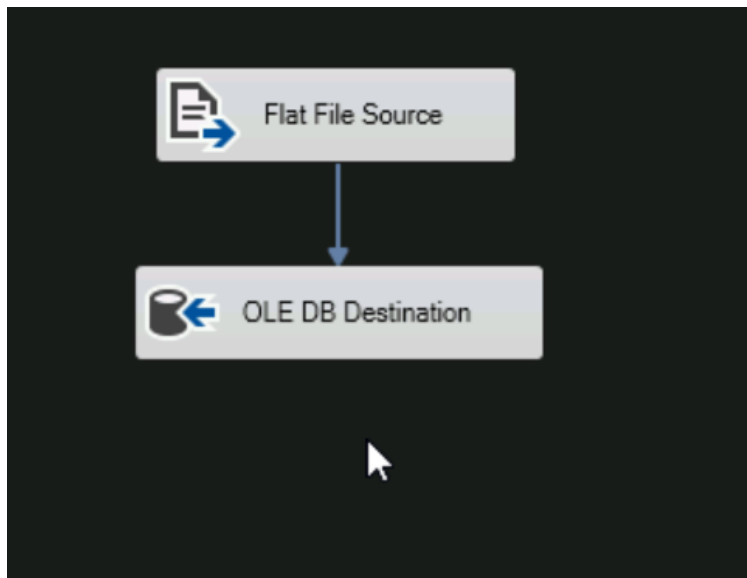
## Assignment – 9 (SSIS)

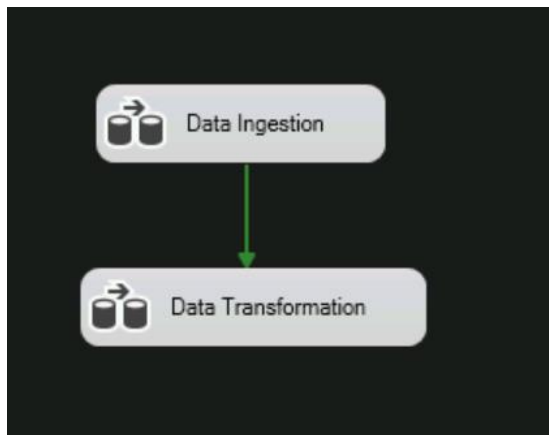
### Task 1: Integration with ETL Data Warehouse (DWH)

Scenario: Your company has a data warehouse designed to consolidate data from various sources for analytical purposes. You need to create an SSIS package that extracts data from a transactional database and loads it into the data warehouse.

Requirements:

1. Create a Connection Manager to connect to the transactional database and the data warehouse.
2. Extract Data from a transactional table (e.g., SalesData) using an OLE DB Source.
3. Transform Data:  
Apply necessary transformations such as data type conversions, data cleansing, and calculations.
4. Load Data into the data warehouse (e.g., FactSales table).





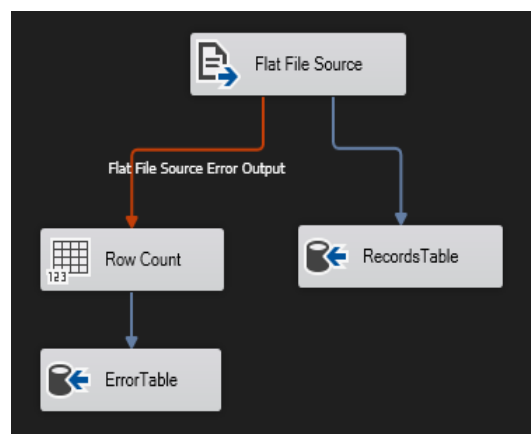
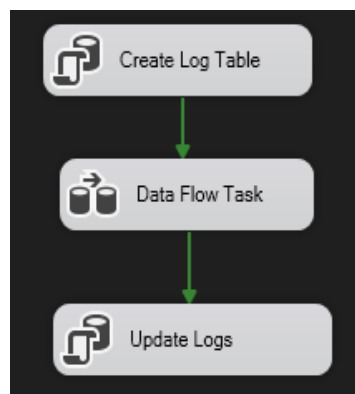
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCTLINE
1	10107	30	95.7	2	2871	2/24/2003 0:00	Shipped	1	2	2003	Motorcycles
2	10121	34	81.35	5	2765.9	5/7/2003 0:00	Shipped	2	5	2003	Motorcycles
3	10134	41	94.74	2	3884.34	7/1/2003 0:00	Shipped	3	7	2003	Motorcycles
4	10145	45	83.26	6	3746.7	8/25/2003 0:00	Shipped	3	8	2003	Motorcycles
5	10159	49	100	14	5205.27	10/10/2003 0:00	Shipped	4	10	2003	Motorcycles
6	10168	36	96.66	1	3479.76	10/28/2003 0:00	Shipped	4	10	2003	Motorcycles
7	10180	29	86.13	9	2497.77	11/11/2003 0:00	Shipped	4	11	2003	Motorcycles
8	10188	48	100	1	5512.32	11/18/2003 0:00	Shipped	4	11	2003	Motorcycles
9	10201	22	98.57	2	2168.54	12/1/2003 0:00	Shipped	4	12	2003	Motorcycles

## Task 2: Data Warehouse Migrations

Scenario: Your organization is migrating its data warehouse from one server to another. You need to create an SSIS package that facilitates this migration.

Requirements:

1. Create Connection Managers for both the source and destination data warehouses.
2. Transfer Data from the source data warehouse to the destination using the Data Flow Task.
3. Ensure Data Integrity:
  - a) Include checks and balances to ensure data is correctly migrated.
  - b) Log the success or failure of the migration process.

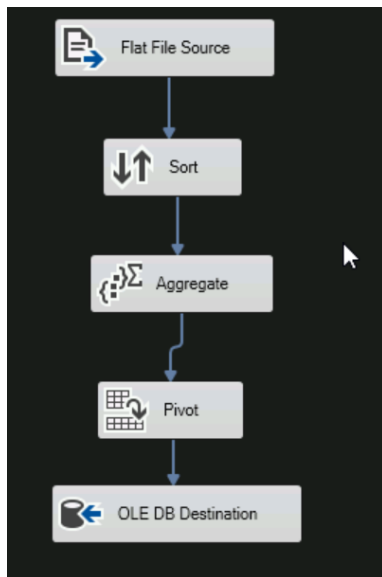


### Task 3: Implementing a Pivot Transformation

Scenario: You have data in a normalized format and need to pivot it for reporting purposes.

Requirements:

1. Extract Data from the source table using an OLE DB Source.
2. Apply a Pivot Transformation to transform the normalized data into a pivoted format.
3. Load the Pivoted Data into a destination table.



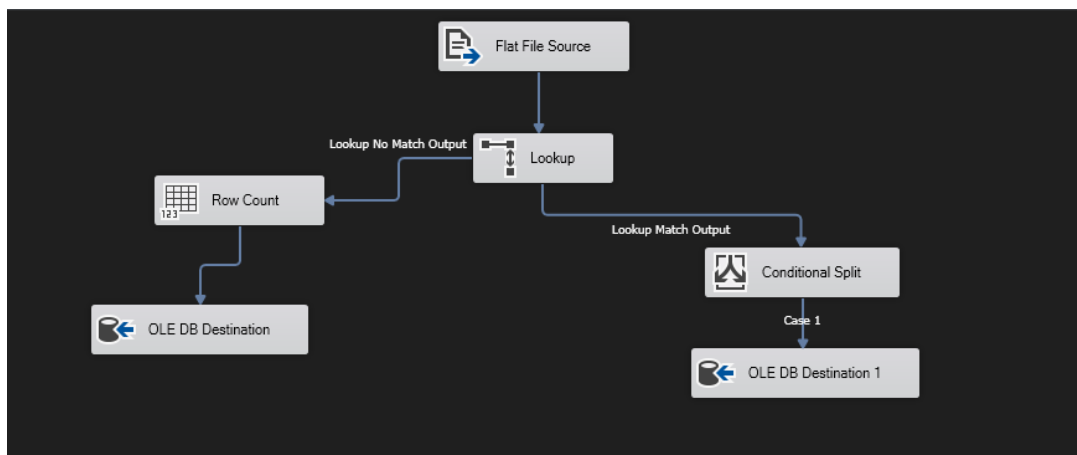
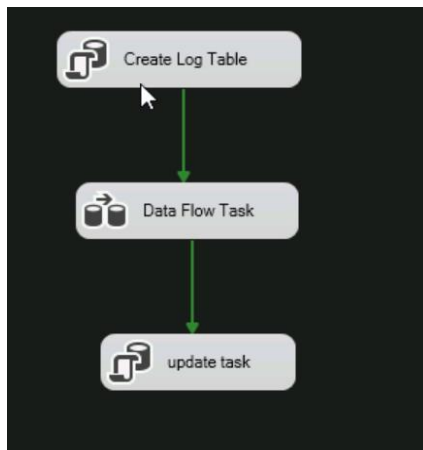
Results		Messages								
	YEAR	Jan	Oct	Nov	Dec	Feb	Mar	Apr	May	June
1	2004	3478.87273365063	3477.5110112796	3618.09969312091	3389.11507568359	3621.15732592206	3673.81660461426	3221.06437110901	3695.11337362753	337
2	2005	3429.73151127979	NULL	NULL	NULL	3692.64101763853	3530.7807525059	4672.02303968157	3815.50883636475	NUI
3	2003	3327.01537772937	3596.77830099758	3479.18129297205	3741.09228428432	3435.02902296113	3490.09800170898	3476.02674286941	3321.95017216123	365

### Task 4: Incremental Load

Scenario: To optimize ETL processes, you need to implement an incremental load to update only the changed data in the data warehouse.

Requirements:

1. Identify Changed Data: Use methods such as timestamps, change data capture using lookup, or checksums.
2. Extract Only the Changed Data from the source.
3. Update the Data Warehouse with the new and changed data only.



100 %

	id	first_name	last_name	gender	email	country
1	18	Brad	Pitt	F	bp@hibu.com	US
2	27	Alan	joy	M	araeburnq@cbc.ca	India
3	38	Home	Lander	M	hmoyne11@huffingtonpost.com	United States
4	40	Tonnie	Montana	M	troyste13@bandcamp.com	China

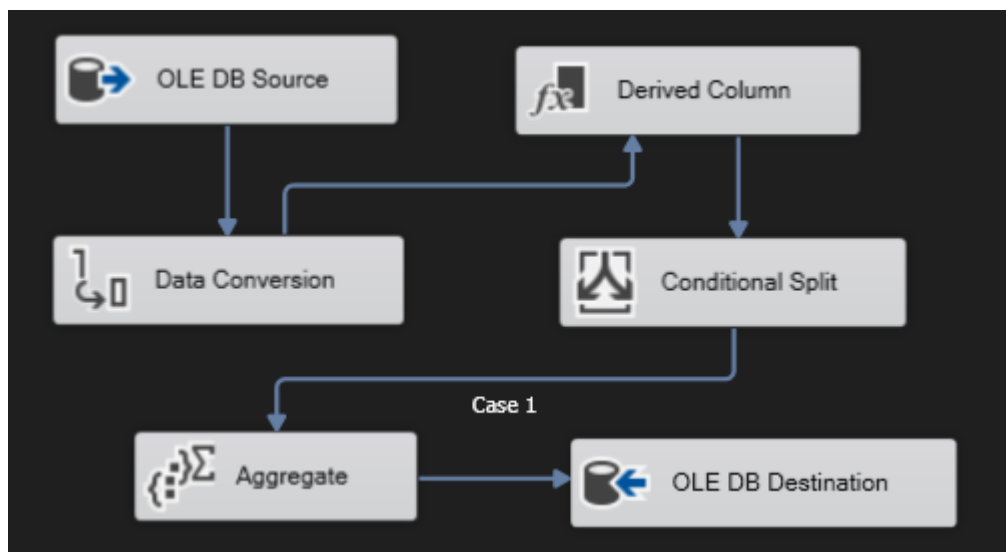
## Task 5: Transformations

Scenario: Your company needs to transform raw data into a format suitable for reporting. You need to perform multiple transformations within an SSIS package.

Requirements:

1. Extract Data from a source table using an OLE DB Source.
2. Apply Transformations such as:
  - Data Conversion
  - Derived Column
  - Conditional Split
  - Aggregate

- Load Transformed Data into a destination table.



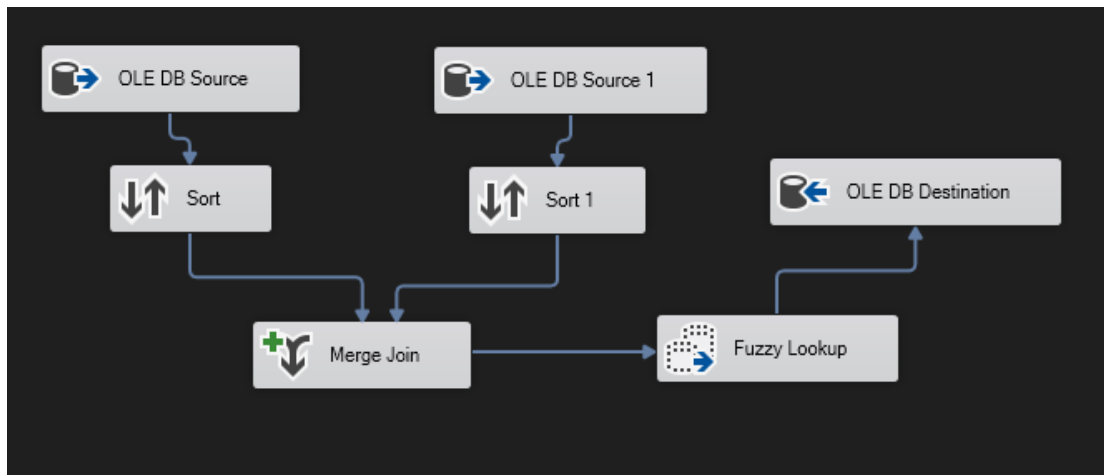
	Data Conversion.YEAR_ID	Data Conversion.QTR_ID	Data Conversion.SALES
1	2003	1	38422.9097900391
2	2003	2	48214.9202880859
3	2003	3	85244.7197265625
4	2003	4	199013.030883789
5	2004	1	90267.0200195313
6	2004	2	130355.429992676
7	2004	3	127949.979309082
8	2004	4	211972.799438477
9	2005	1	135695.610107422
10	2005	2	99251.9196777344

#### Task 6: MERGE & FUZZY LOOKUP

Scenario: You need to merge two datasets and use fuzzy matching to handle potential duplicates.

Requirements:

- Extract Data from two source tables using OLE DB Sources.
- Apply a Merge Join to combine the datasets based on a common key.
- Use Fuzzy Lookup to identify and resolve duplicates in the merged data.
- Load the Cleaned Data into a destination table.



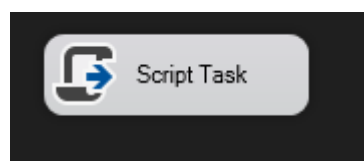
OriginAirportID	city	state	airport_name	DayofMonth	DayOfWeek	Carrier	DestAirportID	DepDelay	ArrDelay	StateFullName	_Similarity	_Confidence	_Similarity_state
14747	Seattle	WA	Seattle/Tacoma International	1	1	AS	14107	-2	5	Washington	1	1	1
13487	Minneapolis	MN	Minneapolis-St Paul International	1	1	DL	13244	-2	-20	Minnesota	1	1	1
14747	Seattle	WA	Seattle/Tacoma International	1	1	VX	12892	-2	-2	Washington	1	1	1
13487	Minneapolis	MN	Minneapolis-St Paul International	1	1	US	14107	-5	-16	Minnesota	1	1	1
14747	Seattle	WA	Seattle/Tacoma International	1	1	OO	14771	-5	2	Washington	1	1	1
13487	Minneapolis	MN	Minneapolis-St Paul International	1	1	AA	13303	-6	9	Minnesota	1	1	1
14747	Seattle	WA	Seattle/Tacoma International	1	1	VX	14771	7	1	Washington	1	1	1
13487	Minneapolis	MN	Minneapolis-St Paul International	1	1	9E	11066	-5	-6	Minnesota	1	1	1
14747	Seattle	WA	Seattle/Tacoma International	1	1	AS	12889	-1	7	Washington	1	1	1
13487	Minneapolis	MN	Minneapolis-St Paul International	1	1	AA	13930	-3	-25	Minnesota	1	1	1
14747	Seattle	WA	Seattle/Tacoma International	1	1	VX	14771	0	4	Washington	1	1	1
13487	Minneapolis	MN	Minneapolis-St Paul International	1	1	DL	14107	-1	-1	Minnesota	1	1	1
14747	Seattle	WA	Seattle/Tacoma International	1	1	VX	12892	-2	-6	Washington	1	1	1

## Task 7: Using Script Task

**Scenario:** You need to perform a complex data transformation that is not supported by the standard SSIS components. A Script Task can be used to achieve this.

Requirements:

1. Add a Script Task to the Control Flow.
2. Write a Script: that performs the required transformation. e.g. Reading data from a file, processing it, and writing the results to a database table.
3. Execute the Script Task within an SSIS package.



airport_id	city	state	airport_name
10135	Allentown/Bethlehem/Easton	PA	Lehigh Valley International
10136	Abilene	TX	Abilene Regional
10140	Albuquerque	NM	Albuquerque International Sunport
10141	Aberdeen	SD	Aberdeen Regional
10146	Albany	GA	Southwest Georgia Regional
10154	Nantucket	MA	Nantucket Memorial
10155	Waco	TX	Waco Regional
10157	Arcata/Eureka	CA	Arcata
10158	Atlantic City	NJ	Atlantic City International
10165	Adak Island	AK	Adak
10170	Kodiak	AK	Kodiak Airport

