

Title: E-commerce Data Integration

Introduction:

E-Commerce wants to improve how it manages data to make better business decisions. This project will build a **data pipeline** in SSIS to collect, clean, and store information for analysis. It includes **data validation** to check accuracy, **bad data handling** to fix or remove errors, **data transformation** to make the data useful, **incremental loading** to update only new data for efficiency, **error and success handling** to track issues and progress, **logging** to record every step, and **scheduling** to automate the process. The final result will be a fully functional SSIS pipeline, a structured database, and documentation explaining everything, helping Shop Smart optimize operations and enhance customer experience.

Data Modeling:

Introduction to Star Schema Data Modeling for EcommerceDW

A star schema is a type of data model used in data warehouses.

- Fact Tables at the center (storing transactions or events)
- Dimension Tables around it (storing descriptive details)

1.Fact Tables (Contain business facts or events)

These are the central tables in the star.

- FactOrders – Stores order transactions (who ordered, when, how much, etc.)
- FactOrderItems – Stores items in each order (product, quantity, supplier, etc.)
- FactReturns – Tracks returned products and refund details
- FactCustomer_Product_Rating – Stores customer reviews and ratings

2.Dimension Tables (Describe the facts in detail)

These are supporting tables around the fact tables.

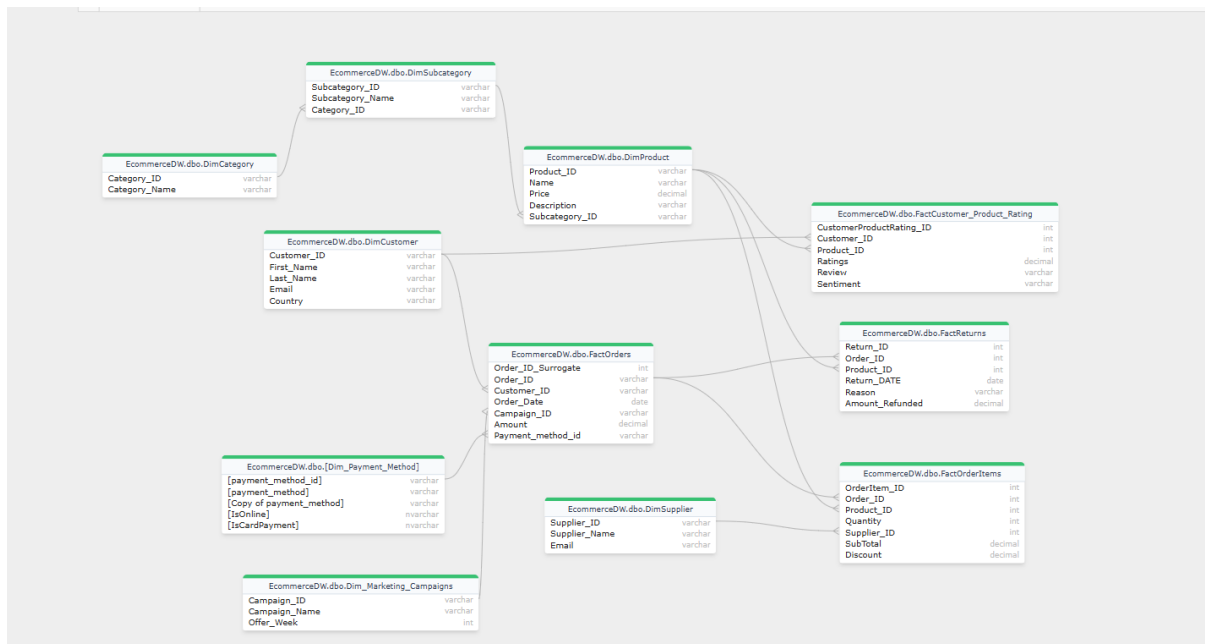
- DimCustomer – Who is buying (name, email, country)
- DimCategory and DimSubcategory – Classify products
- DimProduct – What was purchased (name, price, description)
- DimSupplier – Who supplies the product
- Dim_Payment_Method – How the payment was made
- Dim_Marketing_Campaigns – Which campaign influenced the sale

1.**Simple structure** – Easy to understand and query

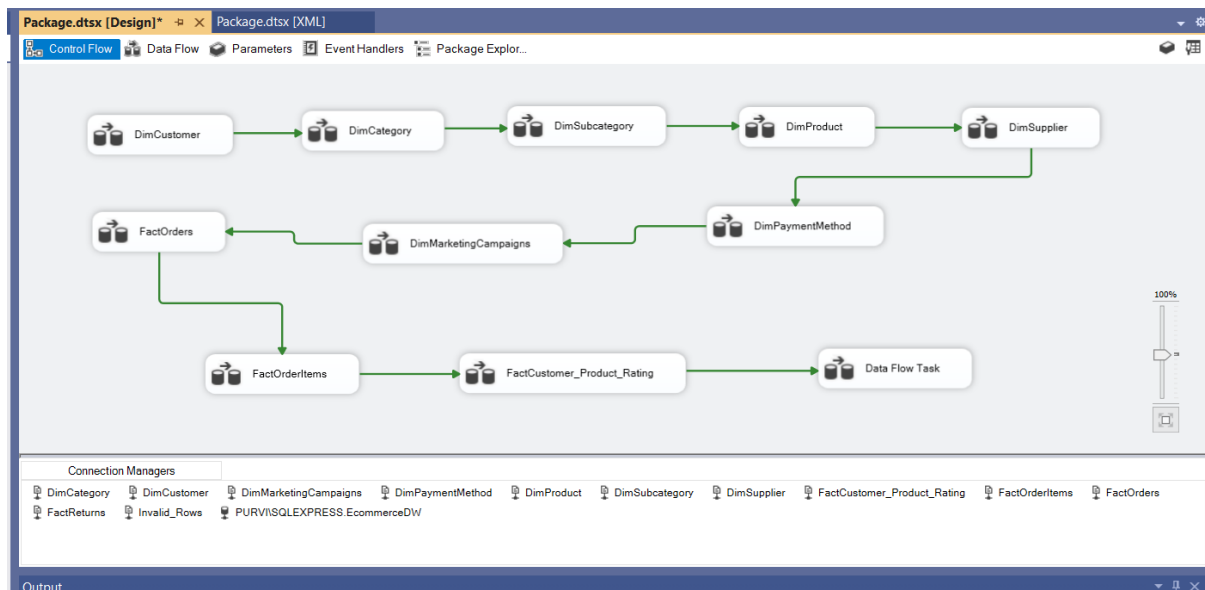
2.**Fast performance** – Optimized for data analysis and reporting

3.**Clear relationships** – Facts linked with descriptive dimensions

ER Diagram:



SSIS ETL Pipeline:



1. DimCustomer ETL Pipeline

Flat File Source: Read Customer CSV file as Source

Derived Column: To Formatting CustomerID, Email and derived full name

Data Conversion : to change data type of fullname and email

Sort: Sorting by Full name

Conditional Split: This Conditional Split filters rows as **Valid** if FormattedCustomerID is not null and not empty, otherwise routes them as **Invalid**.

OLEDB Destination :To store data into Ecommerce DataWarehouse

	FormattedCustomerID	first_name	last_name	email	country	FullName	Email_Domain
1	CUST493	Aaron	Dean	figueroaelizabeth@example.org	Bangladesh	Aaron Dean	example.org
2	CUST290	Aaron	Foster	kochlaura@example.org	Cook Islands	Aaron Foster	example.org
3	CUST488	Aaron	Johnson	yorksteven@example.net	Jordan	Aaron Johnson	example.net
4	CUST078	Aaron	Walker	ucantrell@example.net	Poland	Aaron Walker	example.net
5	CUST393	Adam	Durham	evelynpatel@example.net	Tanzania	Adam Durham	example.net
6	CUST244	Adam	Fernandez	laura18@example.net	Saint Kitts and Nevis	Adam Fernandez	example.net
7	CUST009	Adam	Johnson	pnguyen@example.net	Burundi	Adam Johnson	example.net
8	CUST136	Alexander	Johnson	calhounlisa@example.org	Bhutan	Alexander Johnson	example.org
9	CUST237	Alexandria	Hernandez	renee15@example.net	Denmark	Alexandria Hernandez	example.net
10	CUST171	Alexandria	Thompson	jacksondannny@example.org	Mauritania	Alexandria Thompson	example.org
11	CUST084	Alexis	Thomas	jasmin54@example.com	Puerto Rico	Alexis Thomas	example.com
12	CUST079	Alicia	Anderson	hamiltonhannah@example.org	Australia	Alicia Anderson	example.org
13	CUST395	Alicia	Becker	andrewmartinez@example.net	Falkland Islands (Malvinas)	Alicia Becker	example.net
14	CUST317	Alison	Rodriguez	cindysweeney@example.net	Azerbaijan	Alison Rodriguez	example.net
15	CUST412	Allison	Williams	joseph67@example.org	India	Allison Williams	example.org
16	CUST485	Alyssa	Johnson	shaunrobes@example.org	Maldives	Alyssa Johnson	example.org
17	CUST451	Alyssa	Johnson	shaunrobes@example.org	South Georgia and the South Sandwich Islands	Alyssa Johnson	example.org

2. DimCategory ETL Pipeline

Flat File Source: Read Category CSV file as Source

Derived Column: This Derived Column step checks for null/empty category_name and formats FormattedCategoryID as "CAT" followed by a 3-digit padded number

Conditional Split:

- **ValidRows:** category_name must not be empty after trimming spaces.
- **ValidFormattingCategoryID:** FormattedCategoryID must not be null and its trimmed length must be greater than zero.

Data Conversion:To change data type of category_name

OLEDB Destination: to store data into Dimcategory table in Data Warehouse

	Copy of category_name	Copy of category_id
1	Electronics	CAT001
2	Clothing	CAT002
3	Home & Kitchen	CAT003
4	Books	CAT004
5	Sports & Outdoors	CAT005
6	Toys & Games	CAT006
7	Beauty & Personal Care	CAT007
8	Health & Household	CAT008
9	Automotive	CAT009
10	Tools & Home Improvement	CAT010
11	Office Products	CAT011
12	Grocery	CAT012
13	Pet Supplies	CAT013
14	Musical Instruments	CAT014
15	Movies & TV	CAT015
16	Video Games	CAT016
17	Jewelry	CAT017

3.DimSubcategory ETL Pipeline

Flat File Source: Read Sub Category CSV file as Source

Derived Column: To Formatting Category and Subcategory id

OLEDB Destination: to store data into DimSubcategory table in Data Warehouse

	Derived_subcategory_name	Copy of category_id	Copy of subcategory_id
1	SMARTPHONES	CAT001	SUB001
2	LAPTOPS	CAT001	SUB002
3	HEADPHONES	CAT001	SUB003
4	CAMERAS	CAT001	SUB004
5	WEARABLES	CAT001	SUB005
6	T-SHIRTS	CAT002	SUB006
7	DRESSES	CAT002	SUB007
8	JEANS	CAT002	SUB008
9	SWEATERS	CAT002	SUB009
10	ACTIVEWEAR	CAT002	SUB010
11	FURNITURE	CAT003	SUB011
12	COOKWARE	CAT003	SUB012
13	BEDDING	CAT003	SUB013
14	APPLIANCES	CAT003	SUB014
15	DECOR	CAT003	SUB015
16	FICTION	CAT004	SUB016

4.DimProduct ETL Pipeline

Flat File Source: Read Product CSV file as Source

Data Conversion: To change data type of price and description column

Derived Column: This is a **Derived Column Transformation Editor** (likely from SSIS), used to create or modify data columns. It shows several transformations:

- **PriceNumeric:** Converts 'price' to a double-precision float.
- **PriceWithGST:** Calculates 'price' with 18% GST.
- **name:** Converts 'name' to uppercase.
- **Copy of price:** Creates a copy of the 'price' column.
- **product_id:** Prefixes 'PROD' to a formatted 'product_id'.
- **Copy of subcategory_...:** Prefixes 'SUB' to a formatted 'Copy of subcategory...' value (truncated).

Conditional Split: This is a **Conditional Split Transformation Editor** (likely from SSIS), used to direct input rows to different outputs based on specified conditions.

Here, it splits products into two outputs:

- **ExpensiveProducts:** If PriceNumeric is greater than 750.

- **AffordableProducts:** If PriceNumeric is less than or equal to 750.

OLEDB Destination: to store data into DimExpensiveProduct and DimAffordableProduct table in Data Warehouse

DimAffordableProduct table:

	product_id	name	price	subcategory_id	Copy of description	PriceNumeric	PriceWithGST
1	PROD001	SMARTPHONES - PRODUCT 1	545.54	1	This is a description for Smartphones - Product 1	545.54	643.7372
2	PROD002	SMARTPHONES - PRODUCT 2	59.94	1	This is a description for Smartphones - Product 2	59.94	70.7292
3	PROD003	SMARTPHONES - PRODUCT 3	216.59	1	This is a description for Smartphones - Product 3	216.59	255.5762
4	PROD004	SMARTPHONES - PRODUCT 4	414.26	1	This is a description for Smartphones - Product 4	414.26	488.8268
5	PROD005	SMARTPHONES - PRODUCT 5	431.23	1	This is a description for Smartphones - Product 5	431.23	508.8514
6	PROD006	LAPTOPS - PRODUCT 1	736.51	2	This is a description for Laptops - Product 1	736.51	869.0818
7	PROD007	LAPTOPS - PRODUCT 2	673.74	2	This is a description for Laptops - Product 2	673.74	795.0132
8	PROD008	LAPTOPS - PRODUCT 3	689.21	2	This is a description for Laptops - Product 3	689.21	813.2678
9	PROD010	LAPTOPS - PRODUCT 5	505.61	2	This is a description for Laptops - Product 5	505.61	596.6198
10	PROD011	HEADPHONES - PRODUCT 1	235.61	3	This is a description for Headphones - Product 1	235.61	278.0198
11	PROD013	HEADPHONES - PRODUCT 3	528.03	3	This is a description for Headphones - Product 3	528.03	623.0754
12	PROD016	CAMERAS - PRODUCT 1	615.52	4	This is a description for Cameras - Product 1	615.52	726.3136
13	PROD019	CAMERAS - PRODUCT 4	628.59	4	This is a description for Cameras - Product 4	628.59	741.7362
14	PROD021	WEARABLES - PRODUCT 1	692.04	5	This is a description for Wearables - Product 1	692.04	816.6072
15	PROD022	WEARABLES - PRODUCT 2	98.64	5	This is a description for Wearables - Product 2	98.64	116.3952
16	PROD023	WEARABLES - PRODUCT 3	735.91	5	This is a description for Wearables - Product 3	735.91	868.3738
17	PROD024	WEARABLES - PRODUCT 4	567.42	5	This is a description for Wearables - Product 4	567.42	660.5556

DimExpensiveProduct:

	product_id	name	subcategory_id	Copy of description	PriceNumeric	PriceWithGST
1	PROD009	LAPTOPS - PRODUCT 4	2	This is a description for Laptops - Product 4	917.56	1082.7208
2	PROD012	HEADPHONES - PRODUCT 2	3	This is a description for Headphones - Product 2	869.98	1026.5764
3	PROD014	HEADPHONES - PRODUCT 4	3	This is a description for Headphones - Product 4	993.84	1172.7312
4	PROD015	HEADPHONES - PRODUCT 5	3	This is a description for Headphones - Product 5	766.82	904.8476
5	PROD017	CAMERAS - PRODUCT 2	4	This is a description for Cameras - Product 2	942.05	1111.619
6	PROD018	CAMERAS - PRODUCT 3	4	This is a description for Cameras - Product 3	795.17	938.3006
7	PROD020	CAMERAS - PRODUCT 5	4	This is a description for Cameras - Product 5	944.07	1114.0026
8	PROD027	T-SHIRTS - PRODUCT 2	6	This is a description for T-Shirts - Product 2	901.95	1064.301
9	PROD030	T-SHIRTS - PRODUCT 5	6	This is a description for T-Shirts - Product 5	906.13	1069.2334
10	PROD038	JEANS - PRODUCT 3	8	This is a description for Jeans - Product 3	910.95	1074.921
11	PROD042	SWEATERS - PRODUCT 2	9	This is a description for Sweaters - Product 2	913.46	1077.8828
12	PROD044	SWEATERS - PRODUCT 4	9	This is a description for Sweaters - Product 4	889.65	1049.787
13	PROD048	ACTIVEWEAR - PRODUCT 3	10	This is a description for Activewear - Product 3	826.97	975.8246
14	PROD049	ACTIVEWEAR - PRODUCT 4	10	This is a description for Activewear - Product 4	912.46	1076.7028
15	PROD056	COOKWARE - PRODUCT 1	12	This is a description for Cookware - Product 1	851.3	1004.534
16	PROD061	BEDDING - PRODUCT 1	13	This is a description for Bedding - Product 1	902.94	1065.4692
17	PROD062	BEDDING - PRODUCT 2	13	This is a description for Bedding - Product 2	928.04	1095.0872

4.DimSupplier ETL Pipeline

Flat File Source: Read Supplier CSV file as Source

Data Conversion: To Change data type of email and Supplier name

Derived Column:

- **domain_name:** Extracts the domain from an 'email' address.
- **name_length:** Calculates the length of 'supplier_name'.
- **supplier_id:** Prefixes "SUP" and pads the 'supplier_id' to create a new supplier_id string

Conditional Split: This is an SSIS **Conditional Split Transformation Editor**. It splits data based on the domain_name:

- **ORG_Emails:** Routes rows where domain_name is "example.org" to a table (presumably the ".org table").
- **COM_Emails:** Routes rows where domain_name is "example.com" to another table (presumably the ".com table").

Results Messages

	supplier_id	supplier_name	email	Copy of email	Copy of supplier_name	domain_name	name_length
1	SUP004	Martinez-Vasquez	amber91@example.com	amber91@example.com	Martinez-Vasquez	example.com	16
2	SUP006	Ingram-White	ortizchristopher@example.com	ortizchristopher@example.com	Ingram-White	example.com	12
3	SUP010	Freeman and Sons	desireeballard@example.com	desireeballard@example.com	Freeman and Sons	example.com	16
4	SUP013	"Davis	Underwood and Hernandez",melaniemejia@example.com	Underwood and Hernandez",melaniemejia@example.com	"Davis	example.com	6
5	SUP014	Lamb Group	benjamin10@example.com	benjamin10@example.com	Lamb Group	example.com	10
6	SUP018	Gonzales Group	jnguyen@example.com	jnguyen@example.com	Gonzales Group	example.com	14
7	SUP020	"Crawford	Finley and Collins",coopermatthew@example.com	Finley and Collins",coopermatthew@example.com	"Crawford	example.com	9
8	SUP030	"Dyer	Thompson and White",zduan@example.com	Thompson and White",zduan@example.com	"Dyer	example.com	5

	supplier_id	supplier_name	email	Copy of email	Copy of supplier_name	domain_name	name_length
1	SUP002	Wu Ltd	crystal27@example.org	crystal27@example.org	Wu Ltd	example.org	6
2	SUP005	Hernandez-Cook	ipatterson@example.org	ipatterson@example.org	Hernandez-Cook	example.org	14
3	SUP007	Page-Mccoy	garrettdarrell@example.org	garrettdarrell@example.org	Page-Mccoy	example.org	10
4	SUP009	James Inc	qmontoya@example.org	qmontoya@example.org	James Inc	example.org	9
5	SUP011	"Fitzgerald	Walsh and Williams",ylee@example.org	Walsh and Williams",ylee@example.org	"Fitzgerald	example.org	11
6	SUP015	"Reid	Morgan and Smith",felicia71@exampl...	Morgan and Smith",felicia71@exampl...	"Reid	example.org	5
7	SUP017	"Flores	Fields and Harris",wmiller@example.o...	Fields and Harris",wmiller@example.o...	"Flores	example.org	7
8	SUP022	Bowers-Munoz	xcunry@example.org	xcunry@example.org	Bowers-Munoz	example.org	12

	supplier_id	supplier_name	email	Copy of email	Copy of supplier_name	domain_name	name_length
1	SUP001	"Alexander	Copeland and Smith",christopherpearson@example...	Copeland and Smith",christopherpearson@example...	"Alexander	example.net	10
2	SUP003	"Dixon	Johnston and Gregory",xtapia@example.net	Johnston and Gregory",xtapia@example.net	"Dixon	example.net	6
3	SUP008	"Best	Werner and Clark",dennigarci@example.net	Werner and Clark",dennigarci@example.net	"Best	example.net	5
4	SUP012	Wilcox-Ramir...	andrewhurley@example.net	andrewhurley@example.net	Wilcox-Ramirez	example.net	14
5	SUP016	Gutierrez-Car...	trevorbrown@example.net	trevorbrown@example.net	Gutierrez-Carrillo	example.net	18
6	SUP019	"House	Price and Casey",nathan81@example.net	Price and Casey",nathan81@example.net	"House	example.net	6
7	SUP021	Guzman-Gra...	emily68@example.net	emily68@example.net	Guzman-Graves	example.net	13
8	SUP024	"Howard	Conner and Jones",noblemonique@example.net	Conner and Jones",noblemonique@example.net	"Howard	example.net	7
9	SUP027	Ruiz-Booker	andrew32@example.net	andrew32@example.net	Ruiz-Booker	example.net	11
10	SUP029	Kim LLC	ilawrence@example.net	ilawrence@example.net	Kim LLC	example.net	7
11	SUP031	"Neal	Henry and Curry",jonessamantha@example.net	Henry and Curry",jonessamantha@example.net	"Neal	example.net	5
12	SUP032	"Suarez	Ruiz and Hubbard",ju@example.net	Ruiz and Hubbard",ju@example.net	"Suarez	example.net	7

5. Dim Payment Method ETL Pipeline

Flat File Source: Read Payment Method CSV file as Source

Data Conversion: To change data type of Payment method

Derived Column: This SSIS **Derived Column Transformation Editor** modifies and creates new columns related to payment methods:

- **IsOnline:** Detects if payment_method contains "upi" or "online" (case-insensitive).
- **IsCardPayment:** Detects if payment_method contains "card" (case-insensitive).
- **payment_method:** Converts the existing payment_method to uppercase after trimming.
- **payment_method_id:** Creates a new ID by prefixing "PAY" and padding the existing payment_method_id.

OLEDB Destintation : to Store data into Payment method table into data Warehouse

	payment_method_id	payment_method	Copy of payment_method	IsOnline	IsCardPayment
1	PAY001	CASH	cash	Offline	No
2	PAY002	DEBIT CARD	debit card	Online	Yes
3	PAY003	CREDIT CARD	credit card	Online	Yes
4	PAY004	UPI	upi	Online	No
5	PAY005	NET BANKING	net banking	Online	No

6.DimMarketingCampaigns ETL Pipeline

Flat File Source: Read MarketingCampaign CSV file as Source

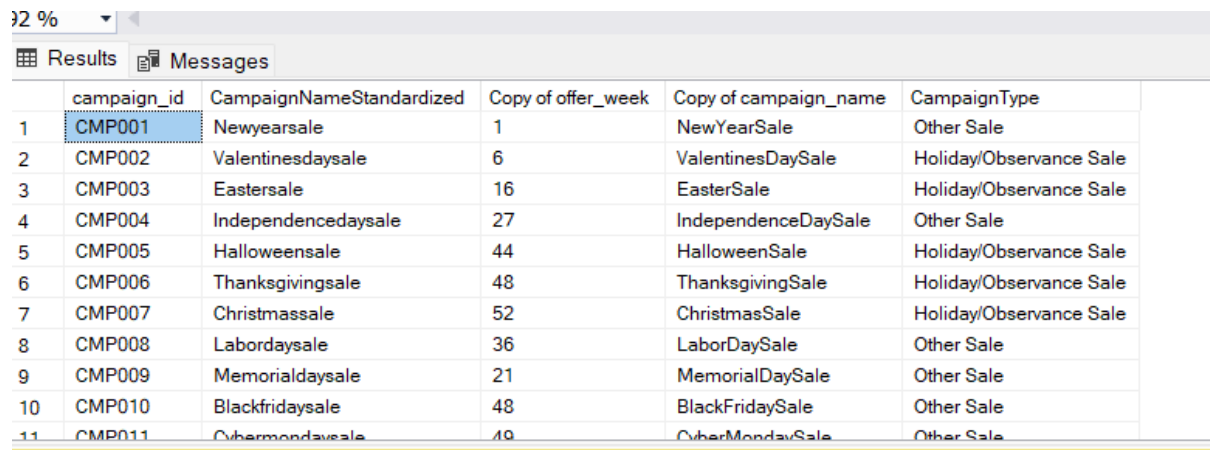
Data Conversion: To change data type of Campaign name and week

Derived Column: This is an SSIS **Derived Column Transformation Editor** with two column operations:

- **CampaignNameStandardize:** Converts the CampaignNameStandardized column to uppercase, effectively standardizing its case.
- **campaign_id:** Creates a new campaign_id by prefixing "CMP" and right-padding the existing campaign_id with zeros.

Script Component: To Write c# Script for find the campaign type and derived new column

OLEDB Destination: To store data into Data Warehouse database



	campaign_id	CampaignNameStandardized	Copy of offer_week	Copy of campaign_name	CampaignType
1	CMP001	Newyearsale	1	NewYearSale	Other Sale
2	CMP002	Valentinesdaysale	6	ValentinesDaySale	Holiday/Observance Sale
3	CMP003	Eastersale	16	EasterSale	Holiday/Observance Sale
4	CMP004	Independencedaysale	27	IndependenceDaySale	Other Sale
5	CMP005	Halloweensale	44	HalloweenSale	Holiday/Observance Sale
6	CMP006	Thanksgivingsale	48	ThanksgivingSale	Holiday/Observance Sale
7	CMP007	Christmassale	52	ChristmasSale	Holiday/Observance Sale
8	CMP008	Labordaysale	36	LaborDaySale	Other Sale
9	CMP009	Memorialdaysale	21	MemorialDaySale	Other Sale
10	CMP010	Blackfridaysale	48	BlackFridaySale	Other Sale
11	CMP011	Cybermondaysale	49	CyberMondaySale	Other Sale

7. FactOrders ETL Pipeline

Flat File Source: Read Order CSV file as Source

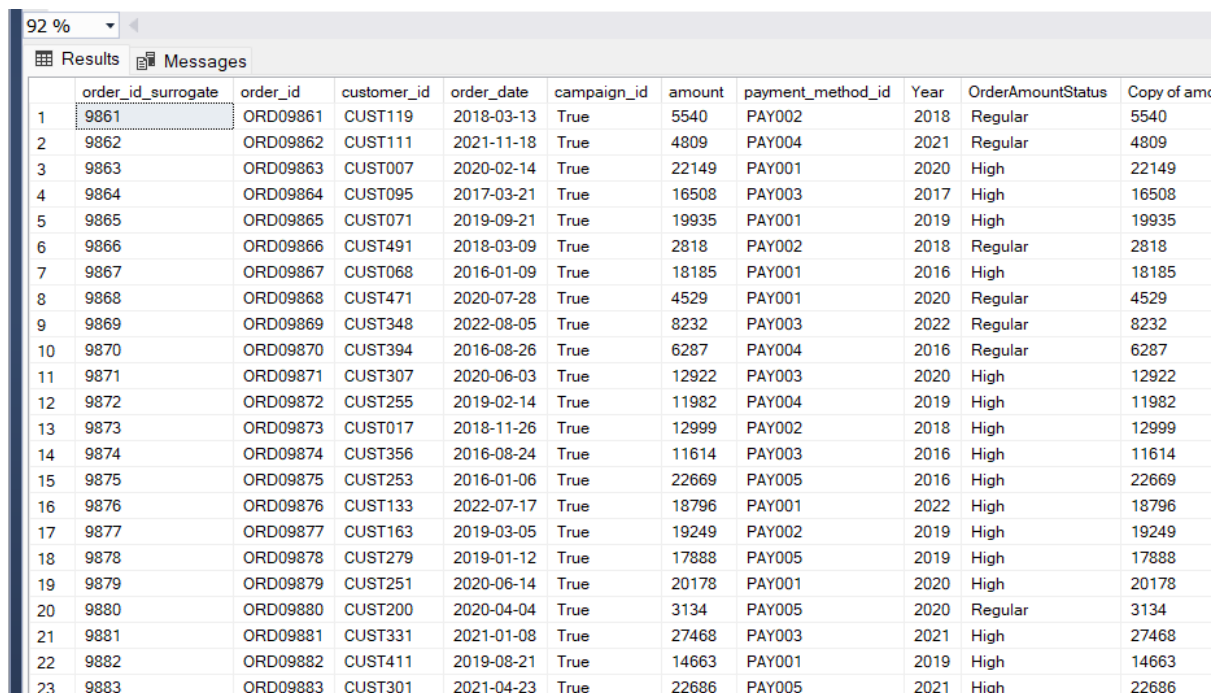
Derived Column: This is an SSIS **Derived Column Transformation Editor** showing several operations:

- **OrderAmountStatus:** Assigns "High" if amount is greater than 10000, otherwise "Regular".
- **order_id:** Prefixes "ORD" and right-pads the order_id.
- **customer_id:** Prefixes "CUST" and right-pads the customer_id.
- **campaign_id:** Prefixes "CMP" and right-pads the campaign_id.
- **payment_method_id:** Prefixes "PAY" and right-pads the payment_method_id.
- **order_date:** Converts order_date to a string in a specific format.

Data Conversion: To change data type of amount

Sort: sorting by order id

OLEDB Destination : to stored data into FactOrder table in data warehouse



	order_id_surrogate	order_id	customer_id	order_date	campaign_id	amount	payment_method_id	Year	OrderAmountStatus	Copy of amr
1	9861	ORD09861	CUST119	2018-03-13	True	5540	PAY002	2018	Regular	5540
2	9862	ORD09862	CUST111	2021-11-18	True	4809	PAY004	2021	Regular	4809
3	9863	ORD09863	CUST007	2020-02-14	True	22149	PAY001	2020	High	22149
4	9864	ORD09864	CUST095	2017-03-21	True	16508	PAY003	2017	High	16508
5	9865	ORD09865	CUST071	2019-09-21	True	19935	PAY001	2019	High	19935
6	9866	ORD09866	CUST491	2018-03-09	True	2818	PAY002	2018	Regular	2818
7	9867	ORD09867	CUST068	2016-01-09	True	18185	PAY001	2016	High	18185
8	9868	ORD09868	CUST471	2020-07-28	True	4529	PAY001	2020	Regular	4529
9	9869	ORD09869	CUST348	2022-08-05	True	8232	PAY003	2022	Regular	8232
10	9870	ORD09870	CUST394	2016-08-26	True	6287	PAY004	2016	Regular	6287
11	9871	ORD09871	CUST307	2020-06-03	True	12922	PAY003	2020	High	12922
12	9872	ORD09872	CUST255	2019-02-14	True	11982	PAY004	2019	High	11982
13	9873	ORD09873	CUST017	2018-11-26	True	12999	PAY002	2018	High	12999
14	9874	ORD09874	CUST356	2016-08-24	True	11614	PAY003	2016	High	11614
15	9875	ORD09875	CUST253	2016-01-06	True	22669	PAY005	2016	High	22669
16	9876	ORD09876	CUST133	2022-07-17	True	18796	PAY001	2022	High	18796
17	9877	ORD09877	CUST163	2019-03-05	True	19249	PAY002	2019	High	19249
18	9878	ORD09878	CUST279	2019-01-12	True	17888	PAY005	2019	High	17888
19	9879	ORD09879	CUST251	2020-06-14	True	20178	PAY001	2020	High	20178
20	9880	ORD09880	CUST200	2020-04-04	True	3134	PAY005	2020	Regular	3134
21	9881	ORD09881	CUST331	2021-01-08	True	27468	PAY003	2021	High	27468
22	9882	ORD09882	CUST411	2019-08-21	True	14663	PAY001	2019	High	14663
23	9883	ORD09883	CUST301	2021-04-23	True	22686	PAY005	2021	High	22686

8.FactOrderItem ETL Pipeline

Flat File Source: Read Order item CSV file as Source

Derived Column: This SSIS **Derived Column Transformation Editor** performs the following operations:

- **netamount:** Calculates subtotal minus discount to get the net amount.
- **product_id:** Prefixes "PROD" and right-pads the product_id.
- **supplier_id:** Prefixes "SUP" and right-pads the supplier_id.
- **order_id:** Prefixes "ORD" and right-pads the order_id with five zeros.

Data Conversion: To change data type of quantity and netamount

Sort: sorting by order id

Lookup: mapping by order id to orderitem id

OLEDB Destination: to store data into FactOrderitem table in data warehouse

Results		Messages						
	orderid_id	order_id	product_id	quantity	supplier_id	subtotal	discount	Net Amount
	218620	ORD39819	PROD150	9	SUP033	4261.23	0.00	4261.23
	218618	ORD39819	PROD474	6	SUP018	761.82	0.00	761.82
	218614	ORD39819	PROD205	9	SUP022	2180.79	0.00	2180.79
	218616	ORD39819	PROD042	2	SUP041	1826.92	0.00	1826.92
	218617	ORD39819	PROD015	5	SUP042	3834.10	0.00	3834.1
	218613	ORD39819	PROD425	1	SUP028	421.59	0.00	421.59
	218621	ORD39820	PROD361	10	SUP029	9255.80	0.00	9255.8
	218622	ORD39820	PROD186	6	SUP025	5785.74	0.00	5785.74
	218623	ORD39820	PROD398	9	SUP017	3236.31	0.00	3236.31
	218625	ORD39821	PROD375	8	SUP004	1998.96	0.18	1639.1472
	218624	ORD39821	PROD459	8	SUP045	4769.76	0.10	4292.784
	218635	ORD39822	PROD017	9	SUP045	8478.45	0.09	7715.3895
	218634	ORD39822	PROD152	5	SUP003	1790.05	0.19	1449.9405
	218633	ORD39822	PROD041	10	SUP046	6944.40	0.08	6388.848
	218632	ORD39822	PROD010	1	SUP003	505.61	0.22	394.3758
	218631	ORD39822	PROD346	5	SUP031	2652.95	0.00	2652.95

9. FactReturns

Flat File Source: Read Return CSV file as Source

Derived Column: This is an SSIS **Derived Column Transformation Editor** with operations to modify existing columns:

- **customer_id:** Prefixes "CUST" and right-pads the customer_id.
- **product_id:** Prefixes "PROD" and right-pads the product_id.
- **review:** Trims leading/trailing spaces from the review column.
- **sentiment:** Trims leading/trailing spaces from the sentiment column.

Data Conversion: To change data type of Rating

Conditional Split: This SSIS **Conditional Split Transformation Editor** categorizes data based on sentiment:

1. **Bad Sentiment:** Directs rows where sentiment is "bad" to a table (e.g., a "Bad Sentiment" table in a data warehouse).
2. **Good Sentiment:** Directs rows where sentiment is "good" to another table (e.g., a "Good Sentiment" table in a data warehouse).

Rows not matching either condition go to the "Conditional Split Default Output" (which could be another table or further processing).

Results		Messages					
	return_id	order_id	product_id	return_date	reason	amount_refunded	Copy of order_id
1	1	ORD28970	PROD167	2016-07-20	RECEIVED WRONG PRODUCT	False	ORD28970
2	10	ORD42742	PROD333	2020-11-23	DEFECTIVE PRODUCT	False	ORD42742
3	100	ORD38981	PROD485	2019-01-19	SIZE DOESN'T FIT	False	ORD38981
4	1000	ORD12408	PROD284	2020-11-03	FOUND A BETTER DEAL	False	ORD12408
5	1001	ORD14140	PROD415	2022-10-04	COLOR DOESN'T MATCH	False	ORD14140
6	1002	ORD15543	PROD187	2016-01-25	SIZE DOESN'T FIT	False	ORD15543
7	1003	ORD16804	PROD070	2022-08-01	FOUND A BETTER DEAL	False	ORD16804
8	1004	ORD04750	PROD484	2019-05-26	DEFECTIVE PRODUCT	False	ORD04750
9	1005	ORD49549	PROD122	2020-03-03	COLOR DOESN'T MATCH	False	ORD49549
10	1006	ORD29456	PROD340	2019-02-25	CHANGED MY MIND	False	ORD29456
11	1007	ORD36582	PROD261	2021-12-19	RECEIVED WRONG PRODUCT	False	ORD36582
12	1008	ORD22868	PROD253	2019-10-23	PRODUCT DAMAGED DURING SHIPPING	False	ORD22868
13	1009	ORD00375	PROD077	2016-01-14	FOUND A BETTER DEAL	False	ORD00375
14	101	ORD22373	PROD182	2019-07-13	RECEIVED WRONG PRODUCT	False	ORD22373
15	1010	ORD21048	PROD104	2021-02-19	COLOR DOESN'T MATCH	False	ORD21048
16	1011	ORD38883	PROD363	2021-11-28	ORDERED BY MISTAKE	False	ORD38883

10.FactCustomerProductRating

Flat File Source: Read Customer Product rating CSV file as Source

Derived Column:This SSIS **Derived Column Transformation Editor** performs the following operations:

- **product_id:** Prefixes "PROD" and right-pads the product_id with zeros.
- **order_id:** Prefixes "ORD" and right-pads the order_id with five zeros.
- **reason:** Trims leading/trailing spaces from the reason column and converts it to

Data Conversion: to change data type of Order id

Sort: sorting by return id

OLEDB Destination: to store data into Fact return table in data ware house

92 %

Results Messages

	customerproductrating_id	customer_id	product_id	ratings	review	sentiment
1	10013	CUST108	PROD460	2	Product didn't meet my expectations at all.	bad
2	10041	CUST261	PROD492	3	Overpriced for the low quality.	bad
3	10042	CUST227	PROD022	2	Overpriced for the low quality.	bad
4	10061	CUST237	PROD145	3	Product didn't meet my expectations at all.	bad
5	10064	CUST380	PROD451	1	Overpriced for the low quality.	bad
6	10071	CUST445	PROD164	3	Overpriced for the low quality.	bad
7	10072	CUST179	PROD173	2	Product didn't meet my expectations at all.	bad
8	10080	CUST263	PROD019	1.5	Product didn't meet my expectations at all.	bad

	customerproductrating_id	customer_id	product_id	ratings	review	sentiment
1	10010	CUST248	PROD383	3.5	Great quality and value for money.	good
2	10017	CUST223	PROD261	4.5	Excellent product! Highly recommended.	good
3	10018	CUST320	PROD212	4	Impressed with the features and performance.	good
4	10021	CUST025	PROD282	4	Impressed with the features and performance.	good
5	10024	CUST431	PROD216	4	Very satisfied with the product. Will buy again.	good
6	10026	CUST392	PROD436	4	Excellent product! Highly recommended.	good
7	10040	CUST365	PROD133	4.5	Great quality and value for money.	good
8	10054	CUST472	PROD437	4	Great quality and value for money.	good
9	10056	CUST363	PROD287	4	Impressed with the features and performance.	good
10	10057	CUST165	PROD081	4	Great quality and value for money.	good
11	10063	CUST341	PROD205	4.5	Couldn't be happier with my purchase.	good
12	10067	CUST445	PROD394	4	Impressed with the features and performance.	good
13	10069	CUST277	PROD034	4.5	Couldn't be happier with my purchase.	good
14	10073	CUST440	PROD058	3.5	Excellent product! Highly recommended.	good
15	10084	CUST441	PROD128	3.5	Very satisfied with the product. Will buy again.	good
16	10089	CUST369	PROD228	4.5	Great quality and value for money.	good