

# DW Design

Ebay E-Commerce



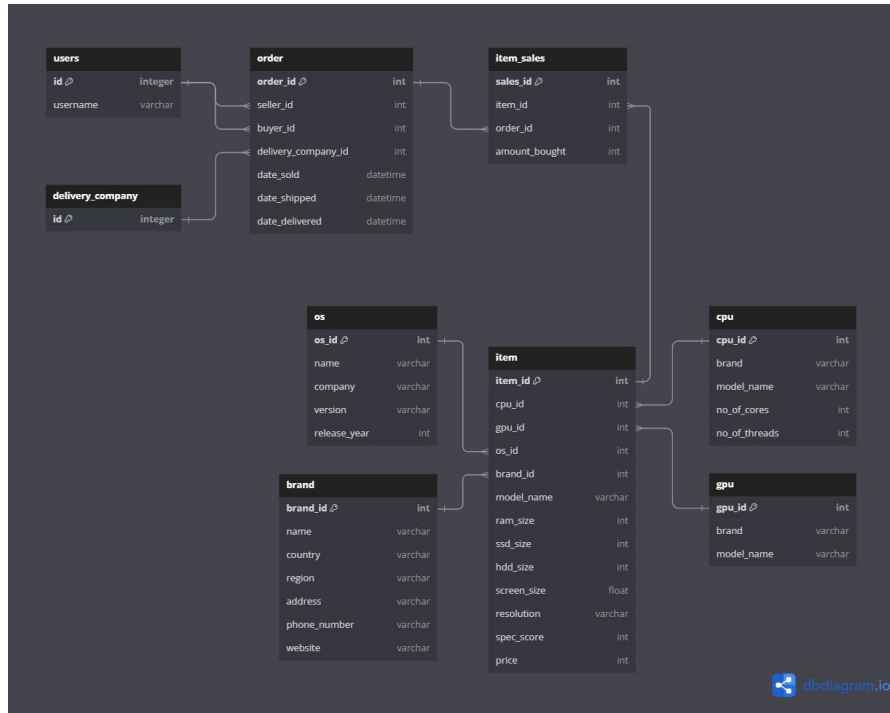


# Introduction



# Data Preparation and Cleaning

- Information generated
  - A list of customers by "<https://www.mockaroo.com/>"
  - Order transaction information from amount and dates by using our personal script.
  - Delivery company by scraping from "<https://www.ebay.com/help/selling/posting-items/setting-postage-options?id=4089>"
- Notebook Information
  - Data of notebook form the year 2022-2023 with detailed information.
    - Model name, processing unit, storage, screen size, memory size, etc.
- Data Assumption
  - Each order can contain multiple items.

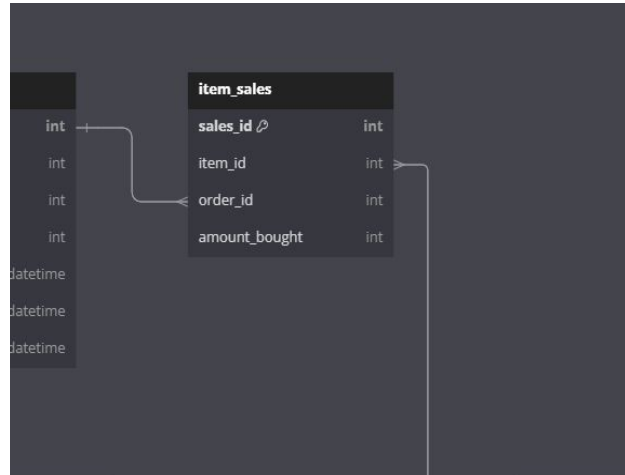


Source Database

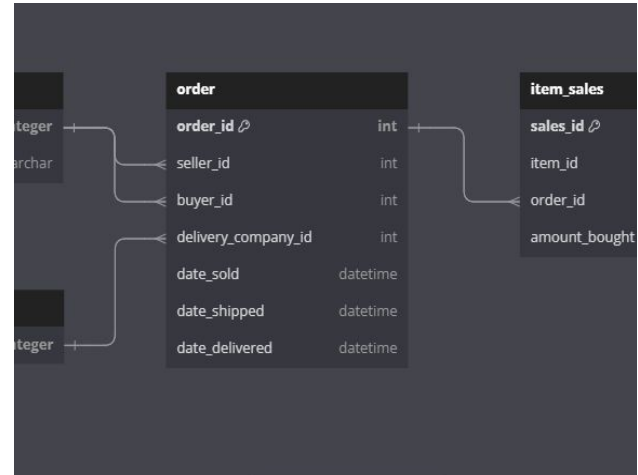


# Identified Business Process

laptop item sales



order lifecycle





# 4 step design



# 4 Steps Design - Item Sale

**Step 1** (Business Process): Tracking item selling

**Step 2** (Granularity): Each row refers to **each sold item** per order each day

**Step 3** (Dimension):

- Laptop\_Dimension
- Date\_Dim
- Time\_Dim
- User\_Dim (2 role playing dimension)

**Step 4** (Fact):

- laptop\_key
- BuyerKey, SellerKey
- DateSoldKey
- TimeSoldKey
- SalesID(DD)
  
- PricePerItem
- NumberOfItemsSold: laptop\_sales.amount\_bought
- TotalPrice: laptop\_price\*amount\_bought

**Table: Item\_Sales\_Fact**

**Columns:**

LaptopKey	int
DateSoldKey	int
BuyerKey	int
SellerKey	int
SalesID	int
PricePerItem	int
NumberOfItemsSold	int
TotalPrice	int
TimeSoldKey	int
<u>SalesFactID</u>	int AI PK



# 4 Steps Design - Order Lifecycle

**Step 1** (Business Process): Tracking order lifecycle in the system

**Step 2** (Granularity): Each row refers to **each order lifecycle**

**Step 3** (Dimension):

- Delivery\_Company\_Dimension
- Date\_Dim (3 role playing dimension)
- Time\_Dim (3 role playing dimension)
- User\_Dim (2 role playing dimension)

**Step 4** (Fact):

- BuyerKey, SellerKey
- DateSoldKey, DateShipmentKey, DateReceivedKey
- TimeSoldKey, TimeShipmentKey, TimeReceivedKey
- DeliveryCompanyKey
- OrderID(DD)
- TotalPrice
- DeliveryCost

**Table: Order\_Lifecycle\_Fact**

**Columns:**

<b><u>BuyerKey</u></b>	int PK
<b><u>SellerKey</u></b>	int PK
<b><u>DeliveryCompanyKey</u></b>	int PK
<b><u>DateSoldKey</u></b>	int PK
<b><u>DateShippedKey</u></b>	int PK
<b><u>DateDeliveredKey</u></b>	int PK
<b><u>TimeSoldKey</u></b>	int PK
<b><u>TimeShippedKey</u></b>	int PK
<b><u>TimeDeliveredKey</u></b>	int PK
OrderID	int
TotalPrice	int
DeliveryCost	int
SoldToShippedLag	int
ShippedToDeliveredLag	int





# SCDs

User\_Dimension: **SCD 2**

DeliveryCompanyDim: **SCD 1**

LaptopDim: **SCD 2**

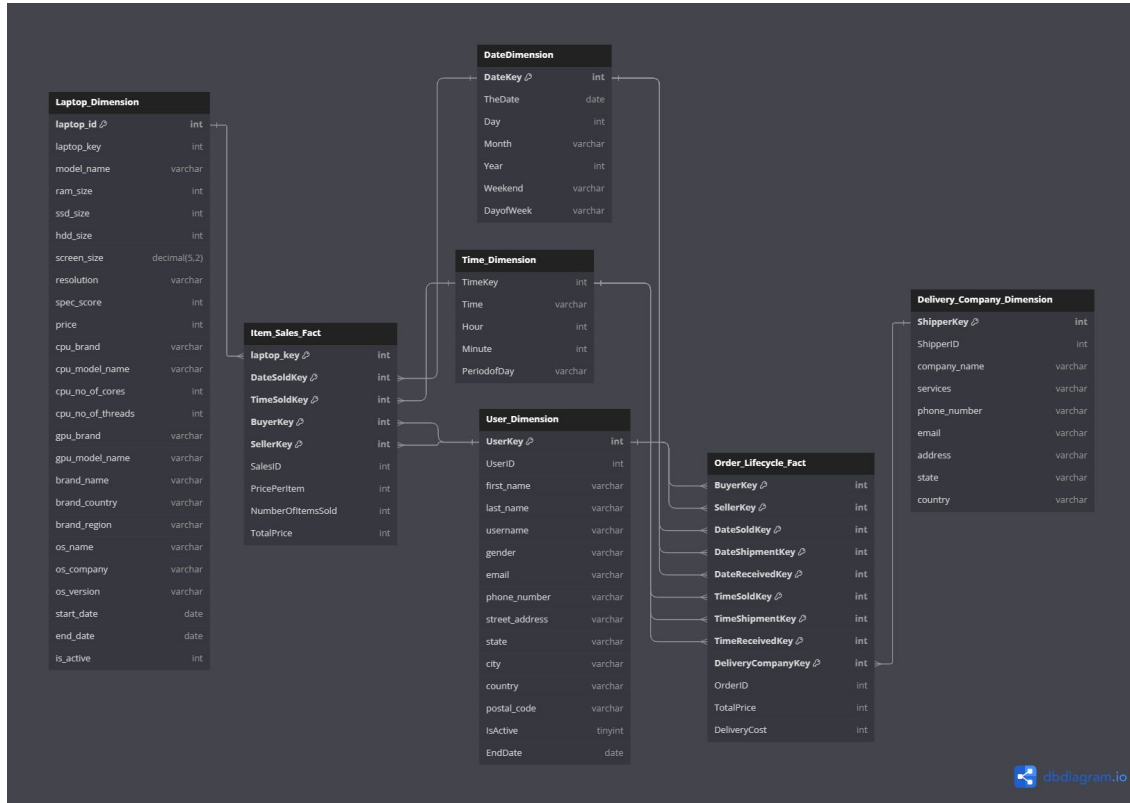
TimeDim: **SCD 0**

DateDim: **SCD 0**



# Conformed Bus Matrix

	DateDim	TimeDim	UserDim	DeliverComp anyDim	LaptopDim
Item Selling	✓	✓	✓		✓
Order Life Cycle	✓	✓	✓	✓	



## Data Warehouse Design

# Implementation on AWS

# Load Dimensions

```
Script info
1 import sys
2 from aws glue.utils import getResolvedOptions
3 from pyspark.context import SparkContext
4 from aws glue.context import GlueContext
5 from aws glue.dynamicframe import DynamicFrame
6 from aws glue.job import Job
7 import pymysql
8 from datetime import date, timedelta
9
10 args = getResolvedOptions(sys.argv, ['JOB_NAME', 'host', 'user', 'pass'])
11
12 sc = SparkContext()
13 glueContext = GlueContext(sc)
14 job = Job(glueContext)
15 job.init(args["JOB_NAME"], args)
16
17 conn = pymysql.connect(
18     host=args["host"],
19     user=args["user"],
20     password=args["pass"],
21 )
22
23 start_date = date(2023, 1, 1)
24 end_date = date(2024, 1, 13)
25 delta = timedelta(days=1)
26
27 while start_date <= end_date:
28     cursor = conn.cursor()
29     command = "INSERT INTO Ebay_dw.Date_Dim_SCD (TheDate, Day, Month, Year, Weekend, DayOfWeek) VALUES (%s, %s, %s, %s, %s, %s);"
30     cursor.execute(command, (
31         start_date.strftime("%Y-%m-%d"),
32         start_date.strftime("%d"),
33         start_date.strftime("%b"),
34         start_date.strftime("%Y"),
35         "Yes" if start_date.weekday() >= 5 else "No",
36         start_date.strftime("%A")
37     ))
38     start_date += delta
39     cursor.close()
40
41 conn.commit()
42 conn.close()
43
44 job.commit()
```

Python Ln 22, Col 1 Errors: 0 Warnings: 0

## AWS Glue - Script - Load Date Dimension

```

import sys
from aws glue. utils import getResolvedOptions
from pyspark. context import SparkContext
from aws glue. context import GlueContext
from aws glue. dynamicframe import DynamicFrame
from aws glue. job import Job
import pymysql
from datetime import datetime, timedelta

args = getResolvedOptions(sys.argv, ['JOB_NAME', 'host', 'user', 'pass'])

sc = SparkContext()
glueContext = GlueContext(sc)
job = Job(glueContext)
job.init(args["JOB_NAME"], args)

conn = pymysql.connect(
    host=args["host"],
    user=args["user"],
    password=args["pass"],
)

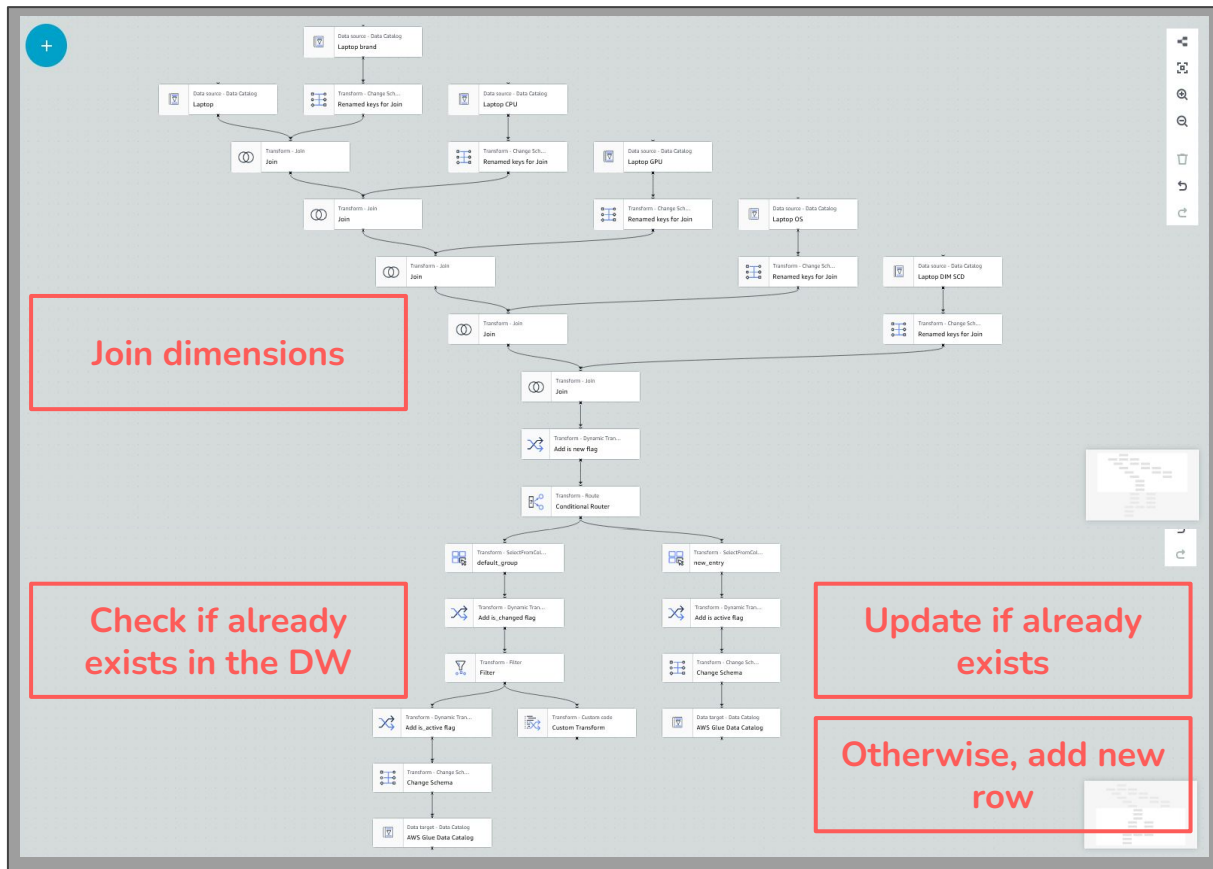
start_time = datetime.strptime("00:00", "%H:%M")
end_time = datetime.strptime("23:59", "%H:%M")
delta = timedelta(minutes=1)

while start_time <= end_time:
    cursor = conn.cursor()
    command = "INSERT INTO Ebay_dw. Time_Dim_SCD (Time, Hour, Minute, PeriodOfDay) VALUES (%s, %s, %s, %s);"
    cursor.execute(command, (
        start_time.strftime("%H:%M"),
        start_time.strftime("%H"),
        start_time.strftime("%M"),
        "Night" if 0 <= start_time.hour < 6 else
        "Morning" if 6 <= start_time.hour < 12 else
        "Afternoon" if 12 <= start_time.hour < 18 else
        "Evening"))
    start_time += delta
    cursor.close()

conn.commit()

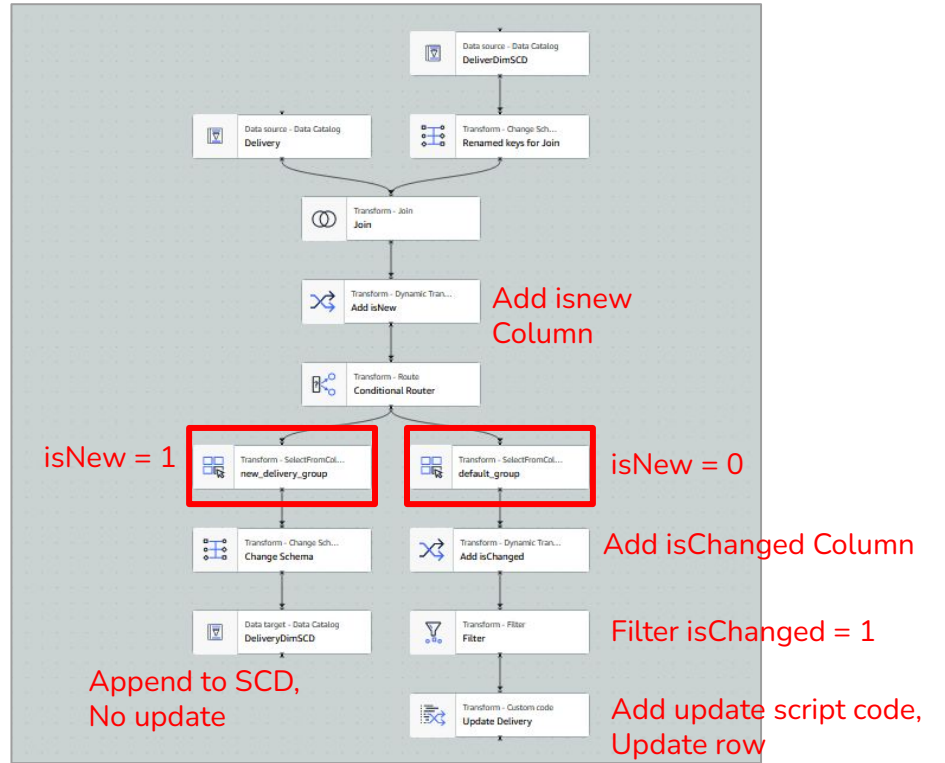
```

## AWS Glue - Script - Load Time Dimension

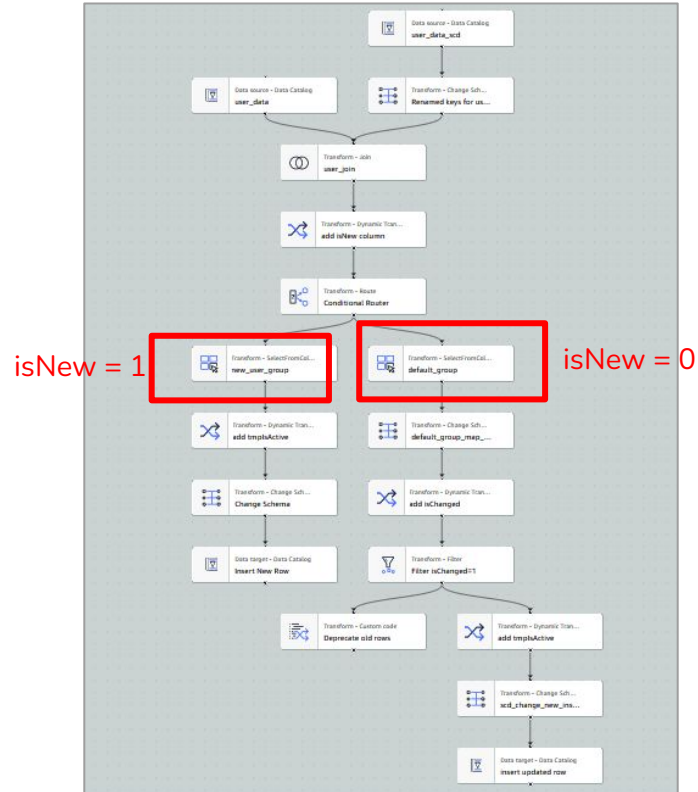


AWS Glue - Visual ETL - Load Laptop Dimension SCD 2





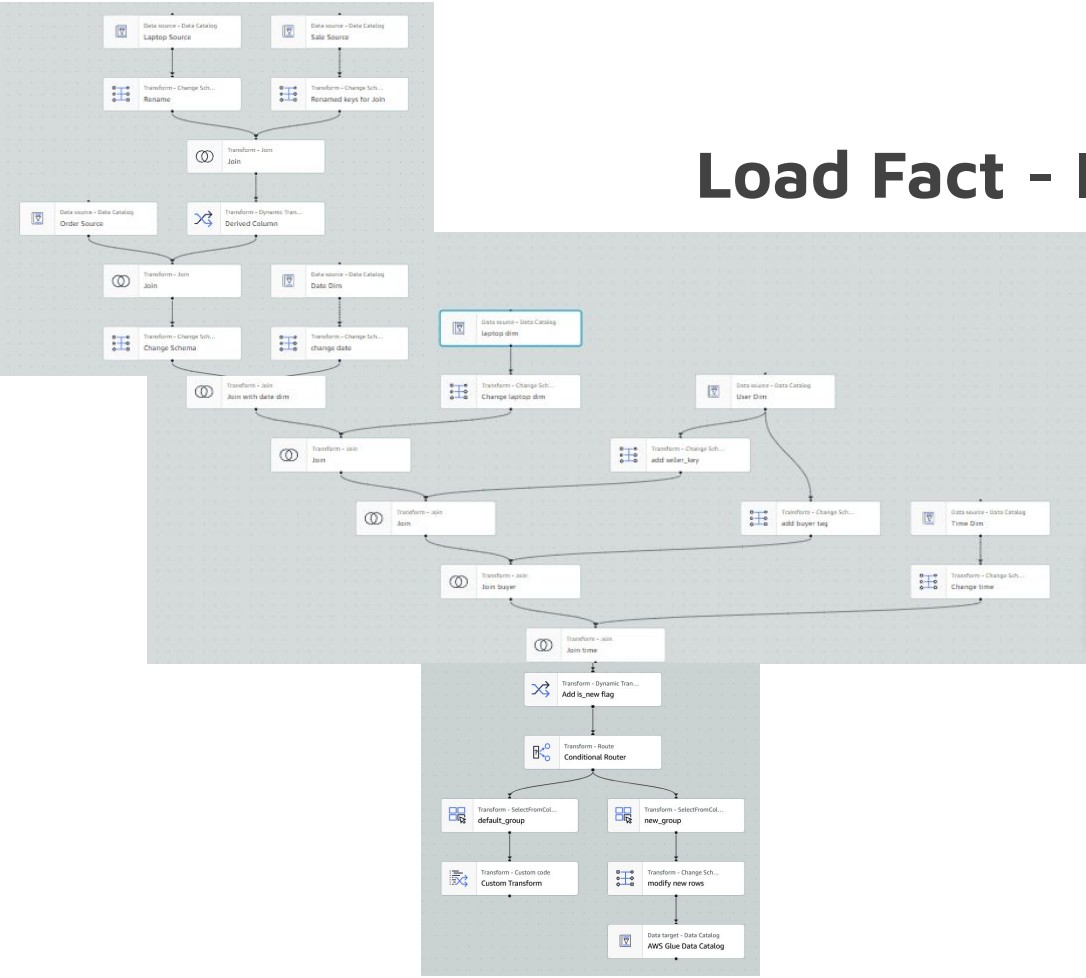
AWS Glue - Visual ETL - Load Delivery Company  
Dimension SCD 1



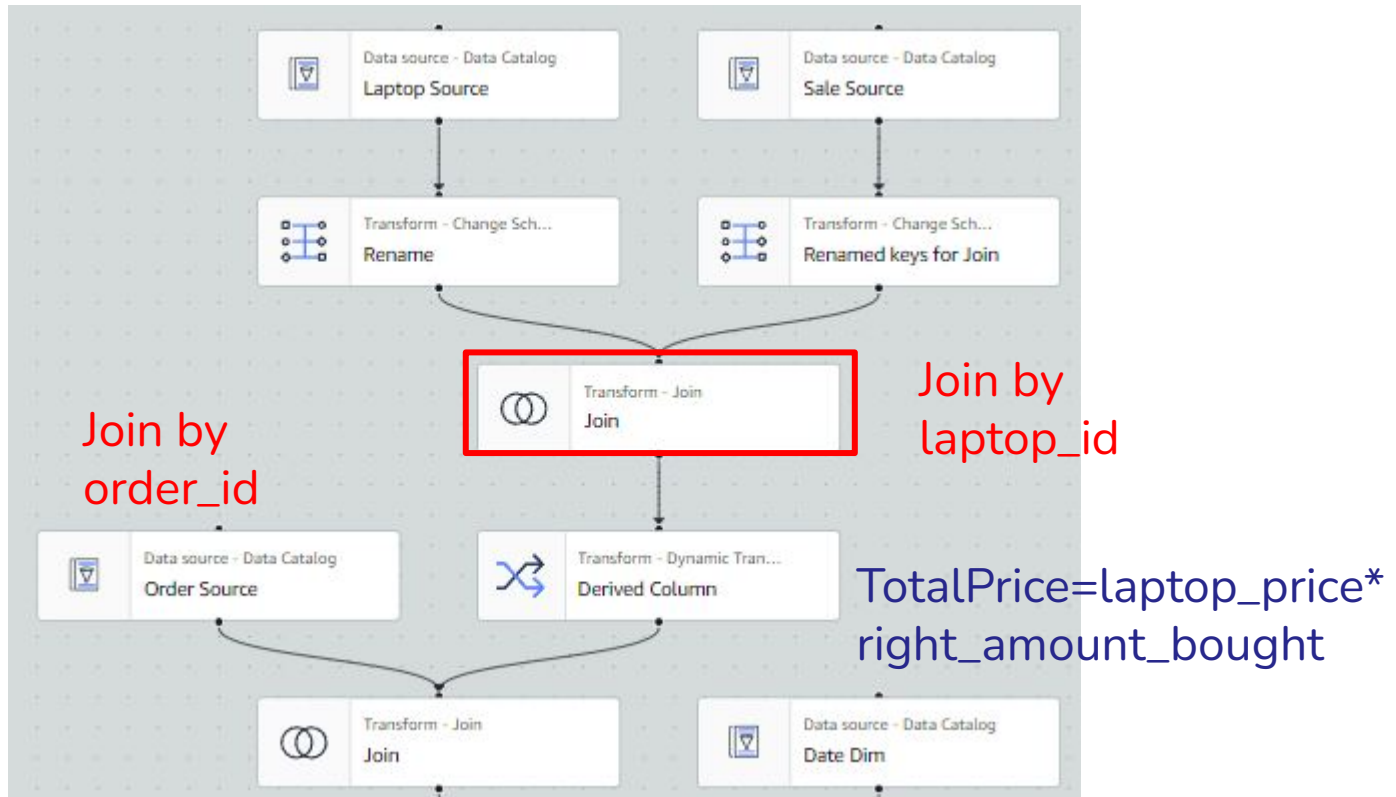
AWS Glue - Visual ETL - Load User Dimension SCD 2

# Load Item Sales Fact

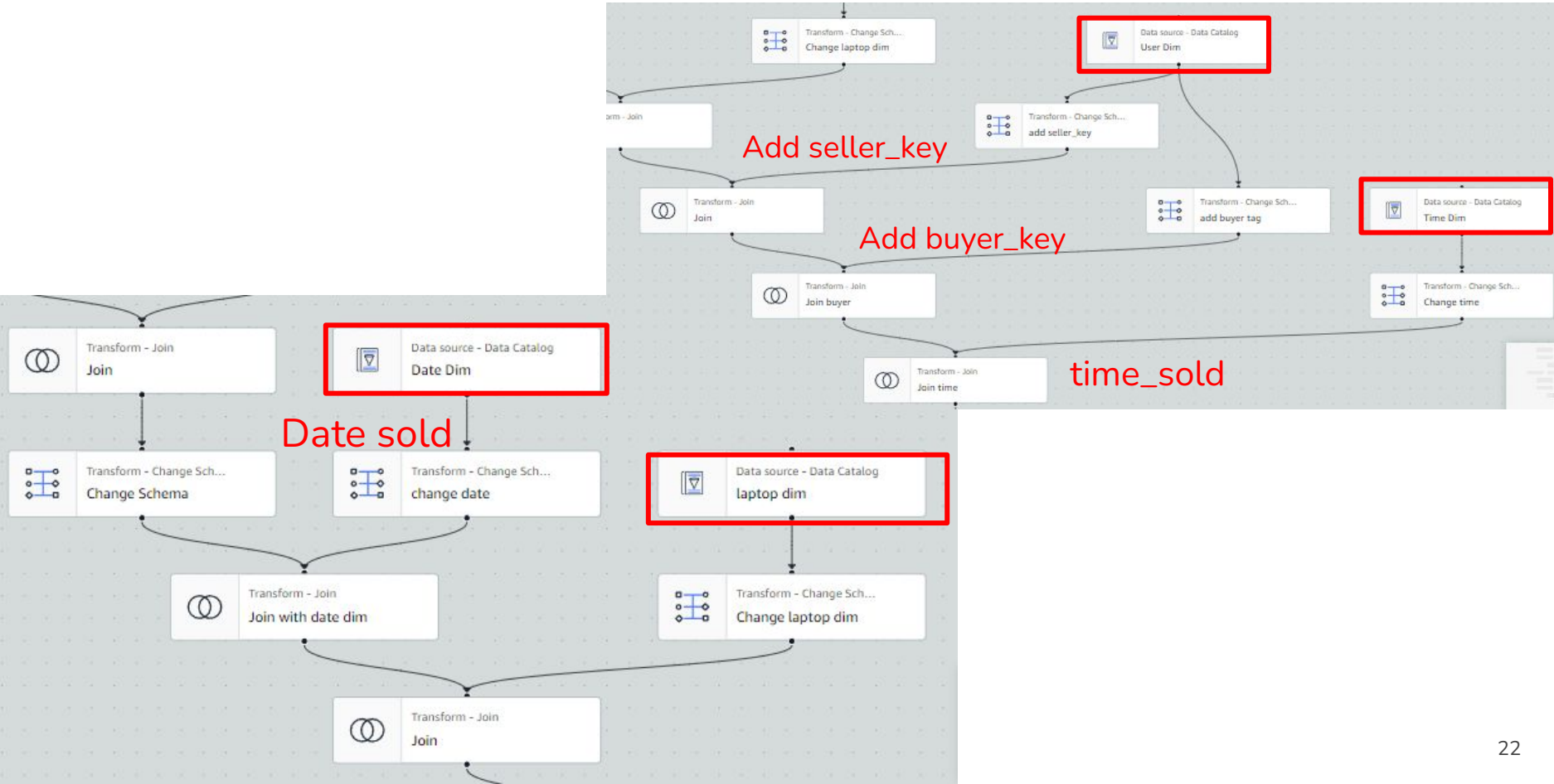
# Load Fact - Item Sales



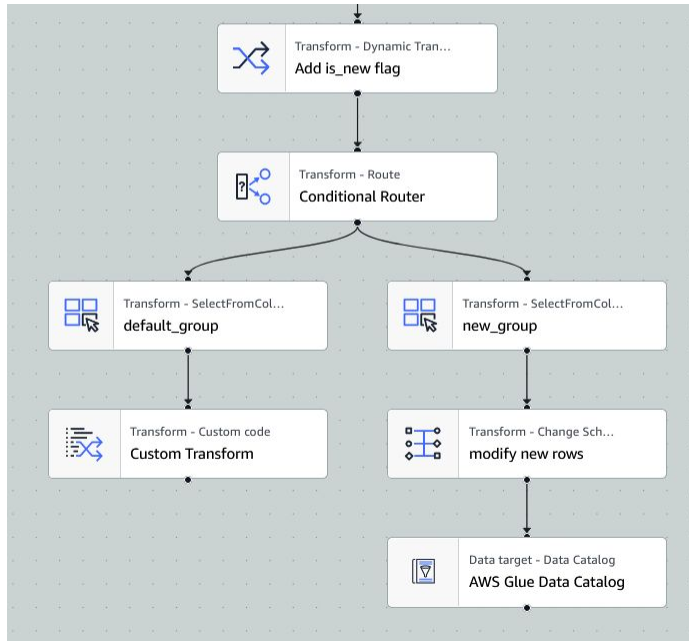
# Load Fact - Item Sales



# Load Fact - Item Sales



# Load Fact - Item Sales

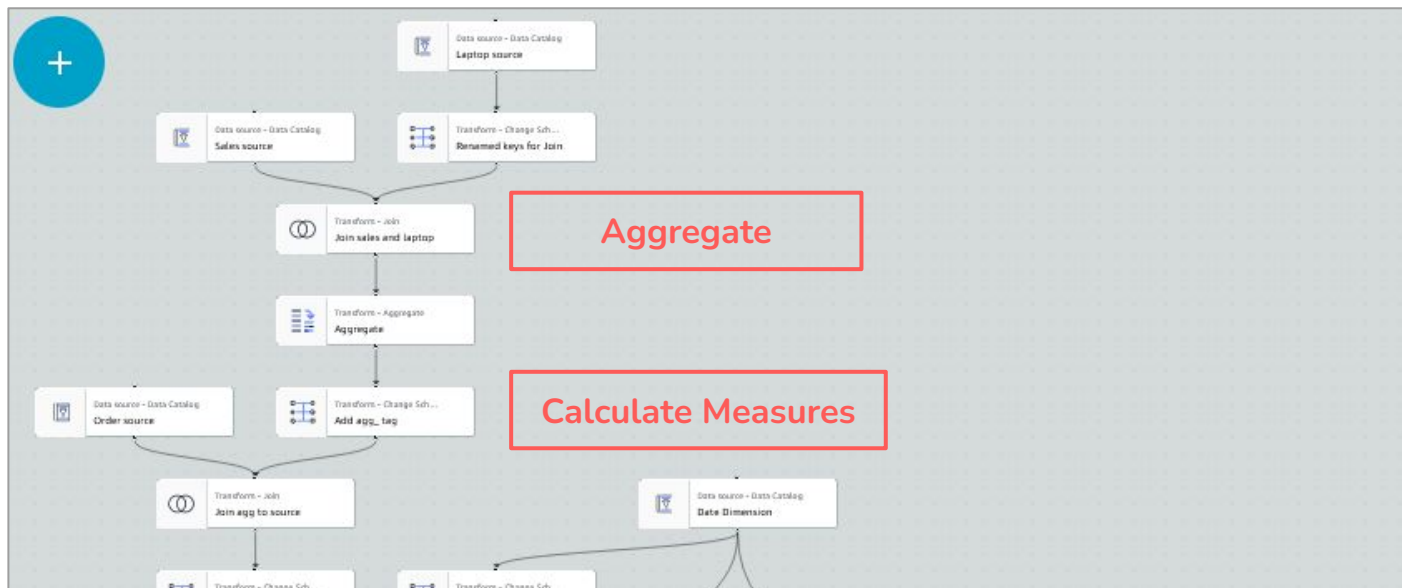


- **datesoldkey:** datekey
- **numberofitemsold:** amount\_bought
- **sellerkey:** seller\_userkey
- **totalprice:** Totalprice
- **buyerkey:** buyer\_userkey
- **laptopkey:** dw\_laptopkey
- **priceperitem:** laptop\_price
- **sale\_id:** sale\_id
- **timesoldkey:** timekey

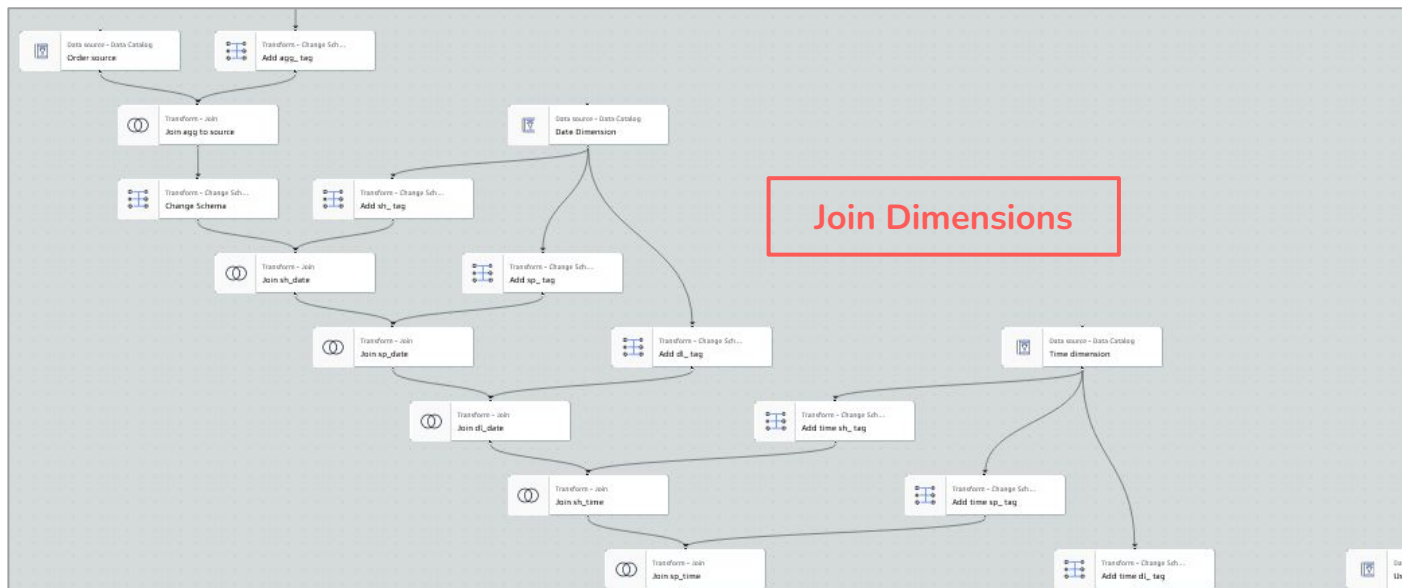
# Load Order Lifecycle Fact



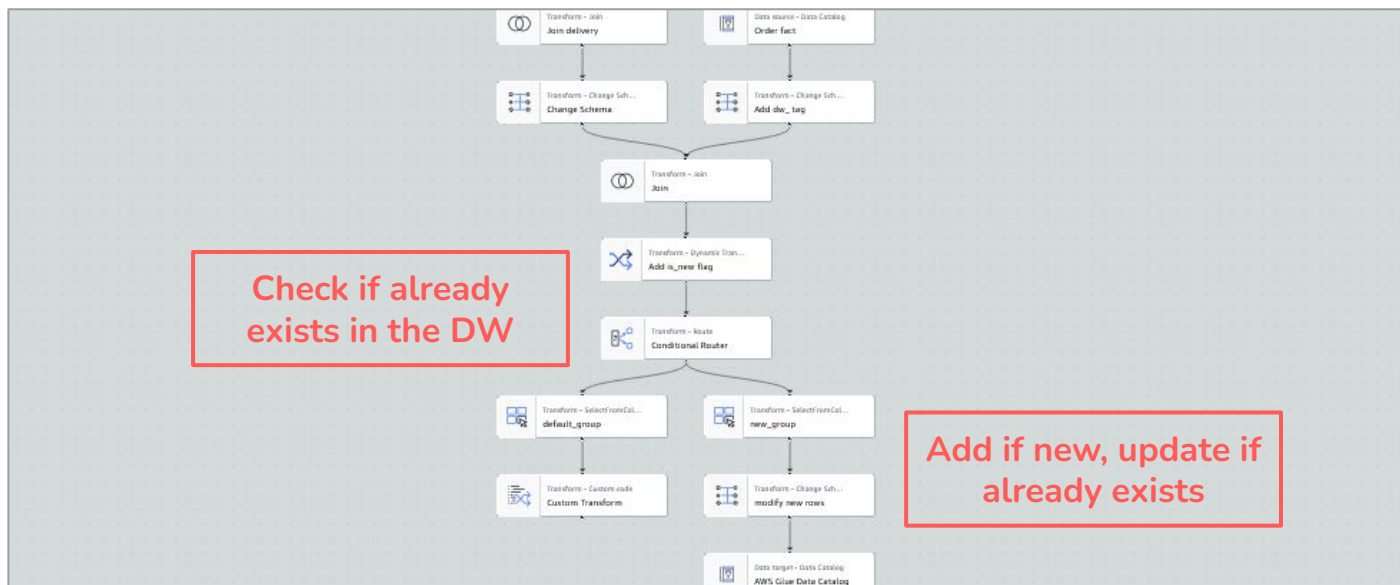
# Load Order Lifecycle Fact: Part I



# Load Order Lifecycle Fact: Part II



# Load Order Lifecycle Fact: Part III



# Results

laptop_id	laptop_key	model_name	ram_size	ssd_size	hdd_size	screen_size	resolution	spec_score	price	cpu_brand	cpu_model_name	cpu_no_of
93	1	HP Pavilion 15-eg2002TU Laptop	16	512	0	15.60	1920 x 1080	65	69990	Intel	12th Gen Core i5	10
37	2	Dell Inspiron 5320 Laptop	16	512	0	13.30	1920 x 1080	65	84918	Intel	12th Gen Core i7	12
101	3	Apple MacBook Pro 16 inch Laptop	32	1000	0	16.20	3456 x 2234	66	349900	Apple	M2 Max 12-core CPU	12
130	4	HP Pavilion 15s-fq5112TU Laptop	16	512	0	15.60	1920 x 1080	55	62490	Intel	12th Gen Core i5	10
110	5	Dell Latitude 3430 Laptop	8	512	0	14.00	1366 x 768	59	108000	Intel	12th Gen Core i5	10
135	6	Acer Nitro 5 AN515-56 Gaming Laptop	16	512	0	15.60	1920 x 1080	73	62990	Intel	11th Gen Core i5	4
126	7	Asus TUF Dash F15 2022 FX517ZC-HN108WS Gaming Laptop	16	1000	0	15.60	1920 x 1080	76	95990	Intel	12th Gen Core i7	12
140	8	Asus TUF Dash F15 2022 FX517ZC-HN035WS Gaming Laptop	8	512	0	15.60	1920 x 1080	72	75990	Intel	12th Gen Core i5	10
299	9	Asus Strix G15 G513RM-HQ271WS Gaming Laptop	16	1000	0	15.60	2560 x 1440	82	124990	AMD	Ryzen 7 6800H	8
245	10	Lenovo IdeaPad Slim 3 82XU017EIN Laptop	8	512	0	15.60	1920 x 1080	65	48990	AMD	AMD Ryzen 5 5500U	6
347	11	MSI Bravo 15 B5DD-043IN Gaming Laptop	8	512	0	15.60	1920 x 1080	69	58500	AMD	Ryzen 5 5600H	6
265	12	Asus ZenBook Pro Duo 15 OLED 2022 UX582ZM-H701WS Laptop	32	1000	0	15.60	3840 x 2160	91	199990	Intel	12th Gen Core i7	12
97	13	HP 14s-fq1092au Laptop	8	512	0	14.00	1920 x 1080	65	46058	AMD	Ryzen 5 5500U	6
204	14	Apple MacBook Pro 16-inch Laptop	16	256	0	15.20	2560 x 1600	64	140000	Apple	Apple M1	8

Laptop Item Dimension Table

ShipperKey	ShipperID	company_name	services	phone_number	email	address	state	country
1	9	Lonestar Overnight (LSO)	Trackable Services	1-800-800-8984	customerservice@lso.com	2800 E Plano Parkway, Suite 300, Plano, Texas...	TX	US
2	8	Spee-Dee Delivery	Trackable Services	320-251-6697	JSmith@speeedelivery.com	4101 Clearwater Road St. Cloud, MN 56301	MN	US
3	6	OnTrac	Trackable Services	1-800-334-5000	webcustomerservice@ontrac.com	1912 Woodford Road, Vienna, Virginia 22182, US	VA	US
4	4	UPS (UPS)	UPS Express Services, UPS Ground Services, UP...	1-888-742-5877	emailinfo@ups.com	55 Glenlake Parkway, NE Atlanta, GA 30328	GA	US
5	5	DHL eCommerce (DHLec)	Trackable Services	+49 228 18296701	AdminContact.Domain@DeutschePost.de	Deutsche Post AG Charles-de-Gaulle-StraÙe 20...	Bonn	DE
6	3	FedEx (FDX)	FedEx Express Services, FedEx Ground/Home D...	1-800-463-3339	janeclarie@fedex.com	FedEx Customer Relations 3875 Airways, Modul...	TN	US
7	7	Lasership	Trackable Services	1-804-414-2590	shipassist@lasership.com	1912 Woodford Rd #2, Vienna, VA 22182, Unit...	VA	US
8	2	United States Postal Service (USPS)	USPS Priority Mail Express, USPS Priority Mail, U...	1-800-275-8777	postalexplorer@usps.gov	Consumer Advocate USPS- 475 L'Enfant Plaza ...	DC	US
9	1	eBay delivery services (eBay)	Trackable Services	1-866-540-3229	jdooe@ebay.com	Ebay, Inc. 2025 Hamilton Avenue San Jose, Cal...	CA	US

Company Dimension Table

	TimeKey	Time	Hour	Minute	PeriodOfDay
▶	1	00:00	0	0	Night
	2	00:01	0	1	Night
	3	00:02	0	2	Night
	4	00:03	0	3	Night
	5	00:04	0	4	Night
	6	00:05	0	5	Night
	7	00:06	0	6	Night
	8	00:07	0	7	Night
	9	00:08	0	8	Night

	DateKey	TheDate	Day	Month	Year	Weekend	DayOfWeek
▶	1	2023-01-01	1	Jan	2023	Yes	Sunday
	2	2023-01-02	2	Jan	2023	No	Monday
	3	2023-01-03	3	Jan	2023	No	Tuesday
	4	2023-01-04	4	Jan	2023	No	Wednesday
	5	2023-01-05	5	Jan	2023	No	Thursday
	6	2023-01-06	6	Jan	2023	No	Friday
	7	2023-01-07	7	Jan	2023	Yes	Saturday
	8	2023-01-08	8	Jan	2023	Yes	Sunday

## Date and Time Dimension Table

UserKey	UserID	first_name	last_name	username	gender	email	phone_number	street_address	state	city	country	postal_code	IsActive	
▶	1	23	Kelly	Blanking	kblankingm	male	kblankingm@twitter.com	6175632527	169 Vermont Lane	Massachusetts	Boston	United States	2163	1
	2	62	Brianna	Macenzyl	bmacenzylp	female	bmacenzylp@archive.org	3048791297	4536 Meadow Ridge Drive	West Virginia	Huntington	United States	25726	1
	3	120	Cory	Vittery	cvittery3b	female	cvittery3b@hubpages.com	9288118335	567 Fuller Crossing	Arizona	Peoria	United States	85383	1
	4	121	Amil	Garred	agarred3c	female	agarred3c@washington.edu	6197585502	92227 Delaware Road	California	San Diego	United States	92105	1
	5	67	Miguel	Machel	mmachel1u	male	mmachel1u@examiner.com	2819018017	15130 Doe Crossing Hill	Texas	Houston	United States	77055	1
	6	6	Beau	Wallworke	bwallworke5	male	bwallworke5@opensource.org	7314458980	466 Parkside Way	Tennessee	Jackson	United States	38308	1
	7	18	Lilan	Marciek	lmarciekh	female	lmarciekh@mac.com	9702708547	458 Doe Crossing Road	Colorado	Fort Collins	United States	80525	1
	8	151	Helli	Edinborough	hedinboroughh6	female	hedinboroughh6@geocities.com	3103620608	1 Bonner Way	California	Garden Grove	United States	92844	1
	9	95	Bogey	Burdoun	bburdoun2m	male	bburdoun2m@home.pl	2025115386	66 Meadow Vale Trail	District of Columbia	Washington	United States	20503	1
	10	19	Gonzales	O'Neill	goneill	male	goneill@arstechnica.com	8164846071	6789 Crowley Lane	Missouri	Kansas City	United States	64193	1
	11	38	Erny	Eglese	eeglese11	male	eeglese11@flickr.com	8139755074	90456 Brentwood Park	Florida	Tampa	United States	33610	1
	12	212	Enrika	Strapp	estrapp5v	female	estrapp5v@disqus.com	8163621580	17 Monterey Street	Missouri	Kansas City	United States	64144	1
	13	115	Dannel	Huggon	dhuggon36	male	dhuggon36@msn.com	7196098571	14332 Roth Parkway	Colorado	Colorado Spri...	United States	80945	1
	14	29	Garrett	Durham	gdurhams	male	gdurhams@cyberchimps.com	9165323224	6737 Roxbury Pass	California	Sacramento	United States	95818	1
	15	46	Markos	MacPhee	mmapchee19	male	mmapchee19@yahoo.co.jp	2159941142	21408 Buena Vista Plaza	Pennsylvania	Philadelphia	United States	19141	1

## User Dimension Table

LaptopKey	DateSoldKey	BuyerKey	SellerKey	SalesID	PricePerItem	NumberOfItemsSold	TotalPrice
883	19	7	354	116	199086	3	597258
189	51	1	5	344	192490	2	384980
532	285	1	11	1733	76990	1	76990
471	46	7	360	315	63990	1	63990
612	285	1	11	1734	48500	1	48500
576	46	7	360	314	56990	1	56990
676	285	1	11	1732	119990	2	239980
334	54	7	408	368	67461	2	134922
73	220	9	1	1314	88990	1	88990
569	320	1	19	1942	49990	1	49990
656	220	9	1	1315	159999	2	319998
971	20	1	70	129	87990	2	175980
930	220	9	1	1316	55850	2	111700
913	232	1	114	1395	159990	1	159990
149	336	9	118	2037	168990	1	168990
938	130	1	246	762	85990	1	85990
590	336	9	118	2038	91958	1	91958
108	363	1	262	2190	187990	5	939950
311	117	9	201	695	29990	1	29990
523	363	1	262	2191	170990	3	512970
264	117	9	201	696	189990	2	379980

Sales Fact Table

BuyerKey	SellerKey	DeliveryCompanyKey	DateSoldKey	DateShippedKey	DateDeliveredKey	TimeSoldKey	TimeShippedKey	TimeDeliveredKey	OrderID	TotalPrice	DeliveryCost
1	5	7	51	61	65	26	1042	888	173	192490	188
1	11	2	285	294	295	182	27	1232	880	245480	62
1	19	9	320	331	336	1413	294	527	981	49990	471
1	70	1	20	23	26	468	1000	190	65	87990	402
1	114	5	232	233	237	767	1417	916	716	159990	320
1	246	8	130	136	141	1000	1105	1317	393	85990	238
1	262	5	363	370	374	1365	860	465	1106	358980	131
1	263	8	116	126	129	956	917	1399	354	209980	312
1	278	5	120	127	132	1263	32	1355	368	415980	476
1	299	2	279	283	286	241	163	667	863	49990	53
1	305	7	39	43	44	1118	673	1053	137	287090	166
1	360	5	202	204	209	364	500	8	613	98000	87
1	360	7	304	307	311	371	252	1276	937	377490	301
1	442	4	346	351	357	1079	903	525	1059	475779	106
2	75	3	106	115	120	762	760	1373	327	175893	201
2	93	1	345	354	356	7	142	63	1058	317480	311
2	94	8	295	301	303	1012	775	564	912	44990	52
2	97	6	214	224	228	910	163	1019	651	42990	496
2	99	7	142	145	150	163	492	451	430	129990	294
2	101	8	319	323	325	156	1087	587	979	229970	201

Order Lifecycle Fact Table





**Thank You**