

EVIDENCIA DE APRENDIZAJE

Proceso de transformación de datos

Elaborado por:

Jeisson Estevens Araque Ramírez.

Docente:

Víctor Hugo Mercado

Institución Universitaria Digital de Antioquia

Ingeniería de Software y Datos

Bases de Datos II

2024

Introducción

En el proceso de crear un data mart nos encontramos con diferentes etapas del proceso como son la extracción la transformación y la carga lo que se denomina ETL por sus siglas en ingles Extract, Transform y Load. El proceso de extracción ya lo realizamos en el trabajo anterior y es hora de continuar con el proceso de transformación de los datos el cual es el corazón del ETL. Aquí es donde los datos extraídos se limpian, se combinan, se enriquecen y se preparan para su análisis. Algunas tareas comunes de transformación incluyen:

Limpieza de datos: Eliminación de duplicados, corrección de errores y normalización.

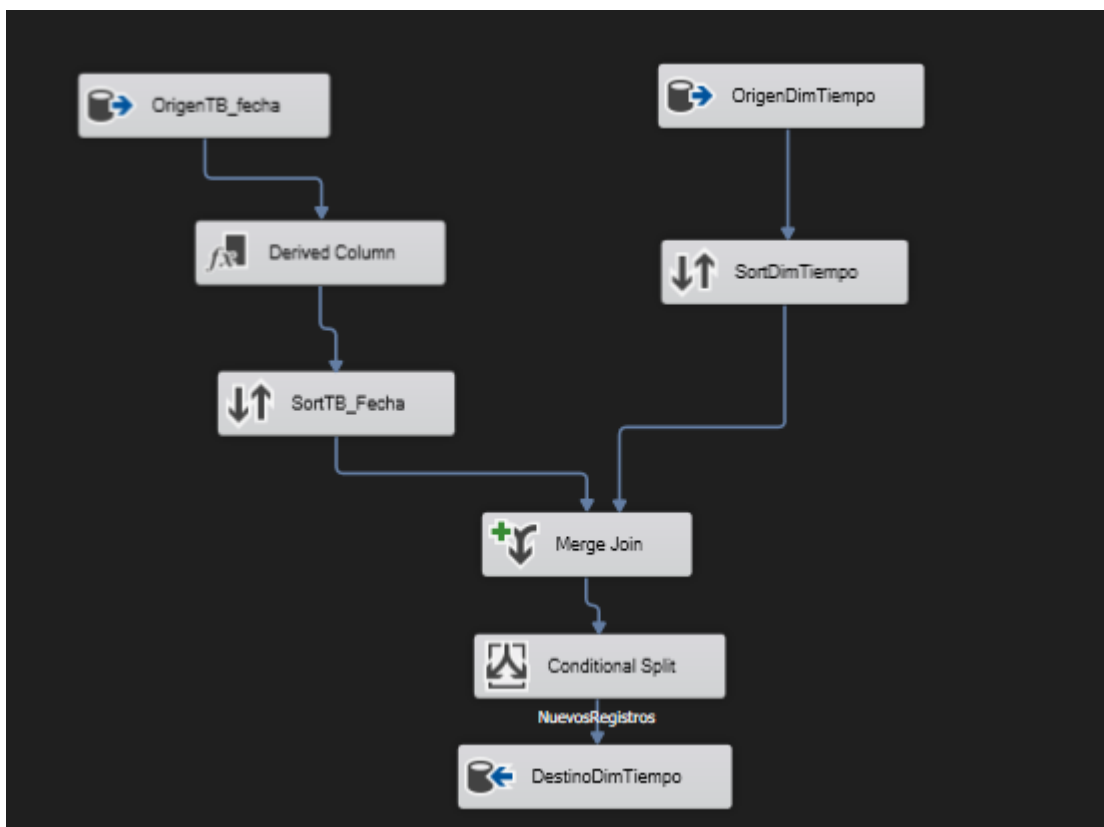
Conversión de formatos: Cambio de tipos de datos, como fechas o números.

Cálculos derivados: Creación de nuevas columnas basadas en fórmulas o reglas de negocio.

La calidad de los datos después de la transformación es crucial para obtener resultados precisos en nuestro data mart.

Transformación de tiempo:

Para construir la dimensión de tiempo fue necesario tomar la tabla de fecha de la base de datos stagin(origen), se eliminan los campos null y usando la herramienta de derived column de visual studio extraemos el año, mes, día, campos definidos para la dimensión de tiempo que cargaremos al modelo estrella.

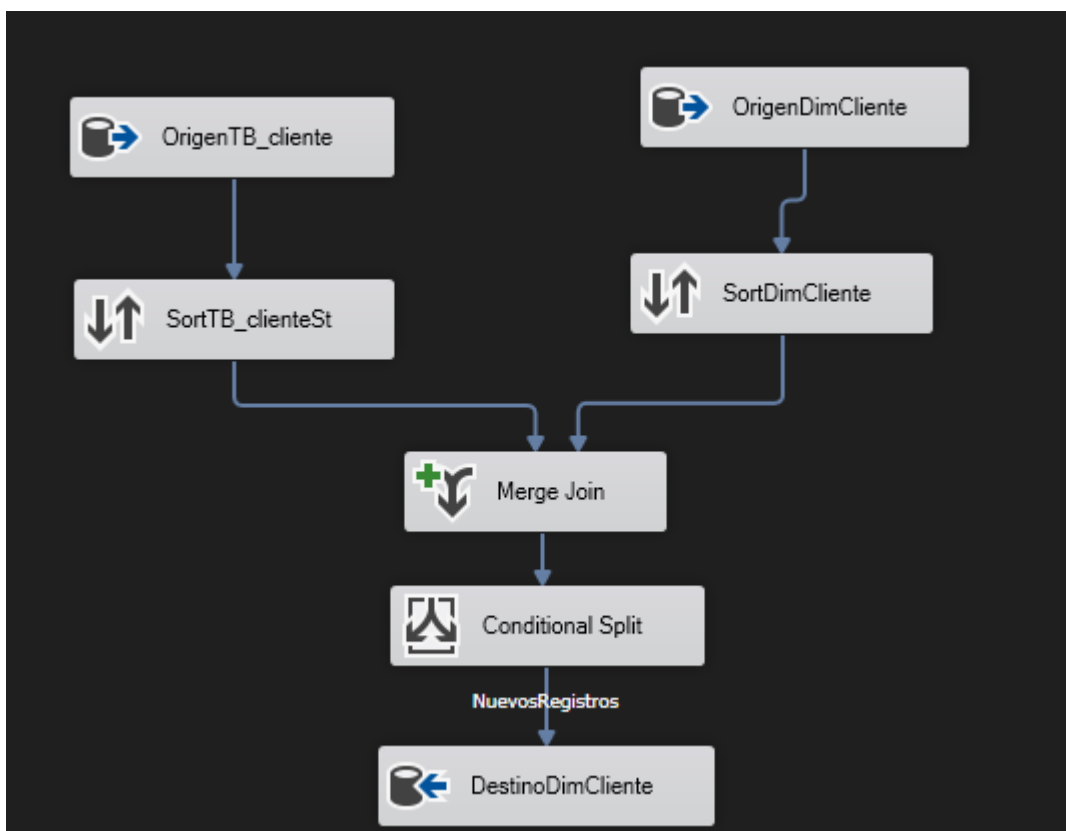


Los datos son llevados de manera correcta a la base de datos:

100 %					
Results Messages					
	ID_fecha	fecha_entrega	Anio	Mes	Dia
1	1	2006-01-19	2006	1	19
2	2	2007-10-26	2007	10	26
3	5	2008-11-14	2008	11	14
4	6	2008-12-28	2008	12	28
5	10	2009-01-15	2009	1	15
6	12	2009-01-11	2009	1	11
7	13	2009-01-15	2009	1	15
8	14	2009-01-11	2009	1	11
9	15	2009-01-07	2009	1	7
10	18	2009-01-09	2009	1	9
11	19	2009-01-13	2009	1	13
12	21	2008-07-25	2008	7	25
13	25	2009-02-20	2009	2	20
14	26	2008-09-01	2008	9	1
15	27	2008-08-31	2008	8	31
16	28	2008-10-04	2008	10	4
17	29	2007-01-27	2007	1	27
18	31	2008-06-28	2008	6	28

Transformación Cliente:

Para crear esta dimensión extraemos los datos de la tabla origen TB_cliente de la base de datos stagin, los campos se pueden cargar así a la dimensión ya que esta tabla no tiene campos nulos ni campos que debamos transformar.

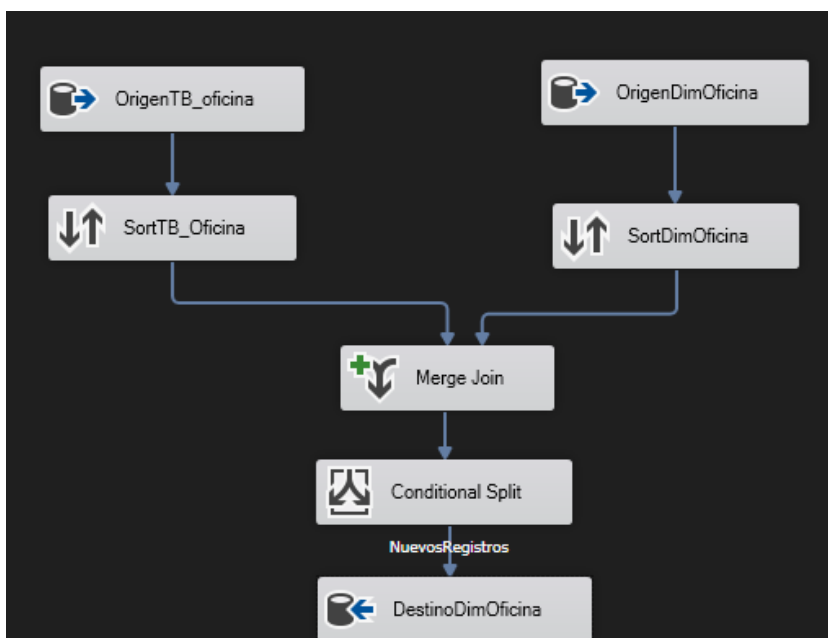


Los datos son llevados de manera correcta a la base de datos:

	ID_cliente	ID_cliente_O	nombre_cliente	telefono	pais	linea_direccion1	ciudad
1	1	1	GoldFish Garden	5556901745	USA	False Street 52 2 A	San Francisco
2	2	2	Gardening Associates	5557410345	USA	Wall-e Avenue	Miami
3	3	3	Gerudo Valley	5552323129	USA	Oaks Avenue n°22	New York
4	4	4	Tendo Garden	55591233210	USA	Null Street n°69	Miami
5	5	5	Lasas S.A.	34916540145	Spain	C/Leganes 15	Fuenlabrada
6	6	6	Beragua	654987321	Spain	C/pintor segundo	Madrid
7	7	7	Club Golf Puerta del hierro	62456810	Spain	C/sinesio delgado	Madrid
8	8	8	Naturagua	689234750	Spain	C/majadahonda	Madrid
9	9	9	DaraDistribuciones	675598001	Spain	C/azores	Madrid
10	10	10	Madridleña de riegos	655983045	Spain	C/Lagañas	Madrid
11	11	11	Lasas S.A.	34916540145	Spain	C/Leganes 15	Fuenlabrada
12	12	12	Camunas Jardines S.L.	34914873241	Spain	C/Virgenes 45	San Lorenzo...
13	13	13	Dardena S.A.	34912453217	Spain	C/Nueva York 74	Madrid
14	14	14	Jardin de Flores	654865643	Spain	C/Oña 34	Madrid
15	15	15	Flores Marivi	666555444	Spain	C/Leganes24	Fuenlabrada
16	16	16	Flowers, S.A	698754159	Spain	C/Luis Salquillo4	Montomes d...
17	17	17	Naturajardin	612343529	Spain	Plaza Magallón 15	Madrid
18	18	18	Golf S.A.	916458762	Spain	C/Estancado	Santa cruz d...
19	19	19	Americh Golf Managem...	964493072	Spain	C/Letardo	Barcelona
20	20	20	Aloha	916485852	Spain	C/Roman 3	Canarias
21	21	21	El Prat	916882323	Spain	Avenida Tibidabo	Barcelona

Transformación Oficina:

Para crear esta dimensión extraemos los datos de la tabla origen TB_oficina de la base de datos stagin, los campos se pueden cargar así a la dimensión ya que esta tabla no tiene campos nulos ni campos que debamos transformar.

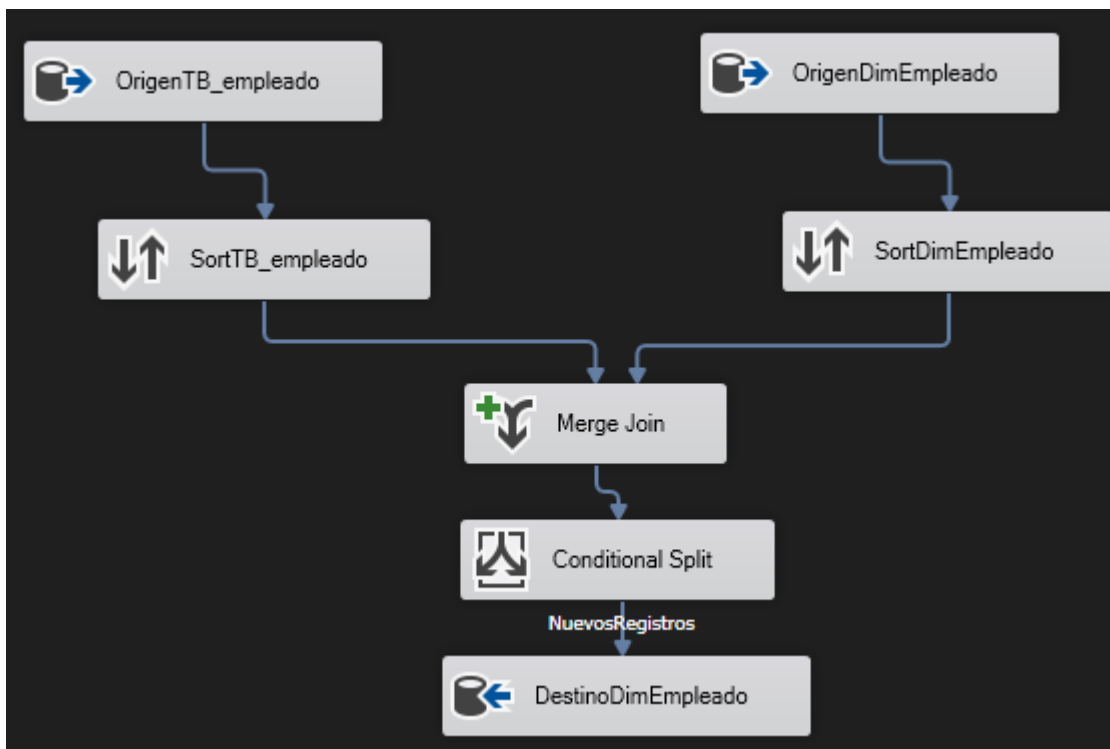


Los datos son llevados de manera correcta a la base de datos:

100 %							
Results Messages							
	ID_oficina	ID_oficina_O	Descripcion	pais	ciudad	telefono	linea_direccion1
1	1	1	BCN-ES	España	Barcelona	+34 93 3561182	Avenida Diagonal, 38
2	2	2	BOS-USA	EEUU	Boston	+1 215 837 0825	1550 Court Place
3	3	3	LON-UK	Inglaterra	Londres	+44 20 78772041	52 Old Broad Street
4	4	4	MAD-ES	España	Madrid	+34 91 7514487	Bulevar Indalecio Prieto, 32
5	5	5	PAR-FR	Francia	Paris	+33 14 723 4404	29 Rue Jouffroy d'abbans
6	6	6	SFC-USA	EEUU	San Francisco	+1 650 219 4782	100 Market Street
7	7	7	SYD-AU	Australia	Sydney	+61 2 9264 2451	5-11 Wentworth Avenue
8	8	8	TAL-ES	España	Talavera de la Reina	+34 925 867231	Francisco Aguirre, 32
9	9	9	TOK-JP	Japón	Tokyo	+81 33 224 5000	4-1 Kioicho

Transformación empleado:

Para crear esta dimensión extraemos los datos de la tabla origen TB_empleado de la base de datos stagin, los campos se pueden cargar así a la dimensión ya que esta tabla no tiene campos nulos ni campos que debamos transformar.

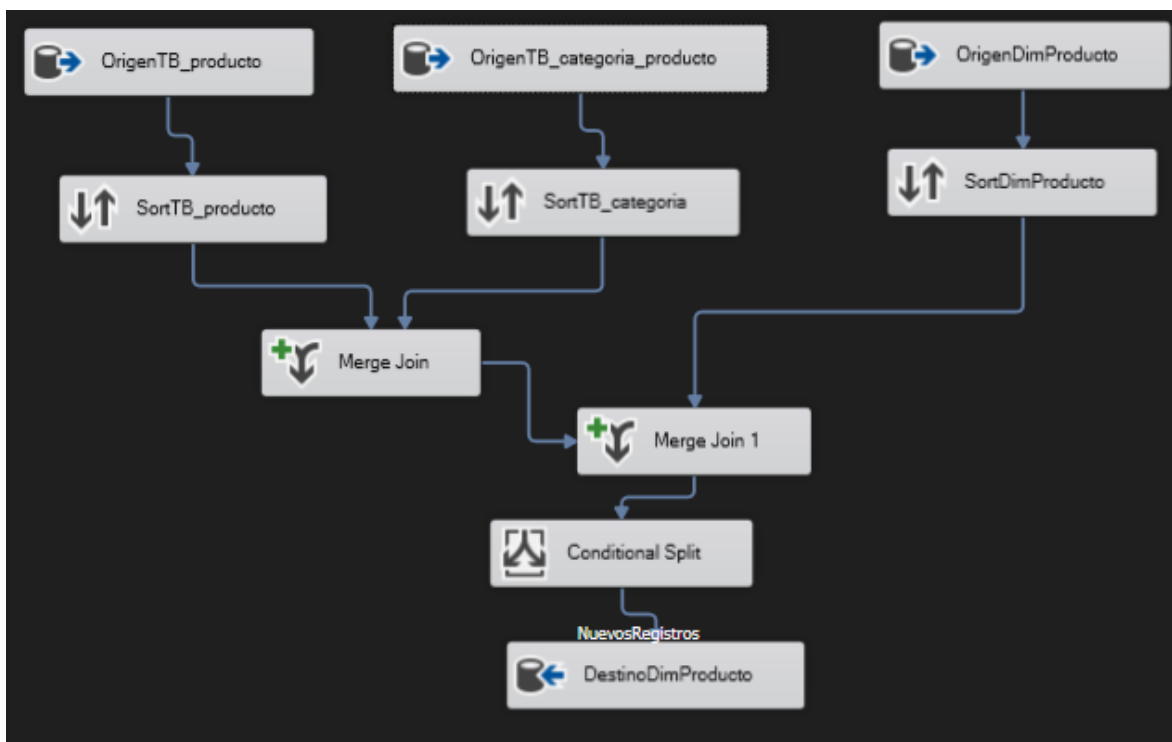


Los datos son llevados de manera correcta a la base de datos:

100 %						
Results Messages						
	ID_empleado	ID_empleado_O	nombre	apellido1	puesto	extension
1	1	1	Marcos	Magaña	Director General	3897
2	2	2	Ruben	López	Subdirector Marketing	2899
3	3	3	Alberto	Soria	Subdirector Ventas	2837
4	4	4	Maria	Solis	Secretaria	2847
5	5	5	Felipe	Rosas	Representante Ventas	2844
6	6	6	Juan Carlos	Ortiz	Representante Ventas	2845
7	7	7	Carlos	Soria	Director Oficina	2444
8	8	8	Mariano	López	Representante Ventas	2442
9	9	9	Lucio	Campoamor	Representante Ventas	2442
10	10	10	Hilario	Rodriguez	Representante Ventas	2444
11	11	11	Emmanuel	Magaña	Director Oficina	2518
12	12	12	José Manuel	Martinez	Representante Ventas	2519
13	13	13	David	Palma	Representante Ventas	2519
14	14	14	Oscar	Palma	Representante Ventas	2519
15	15	15	Francois	Fignon	Director Oficina	9981
16	16	16	Lionel	Narvaez	Representante Ventas	9982
17	17	17	Laurent	Serra	Representante Ventas	9982
18	18	18	Michael	Bolton	Director Oficina	7454
19	19	19	Walter Santiago	Sanchez	Representante Ventas	7454

Transformación Producto:

Para crear esta dimensión extraemos los datos de la tabla origen TB_Producto de la base de datos staging se hace un join con la tabla de categoría_producto para extraer los campos desc_categoria y descripción_texto los cuales previamente se definieron para llevarlos al modelo estrella ya que aportan valor al estudio de business intelligence.

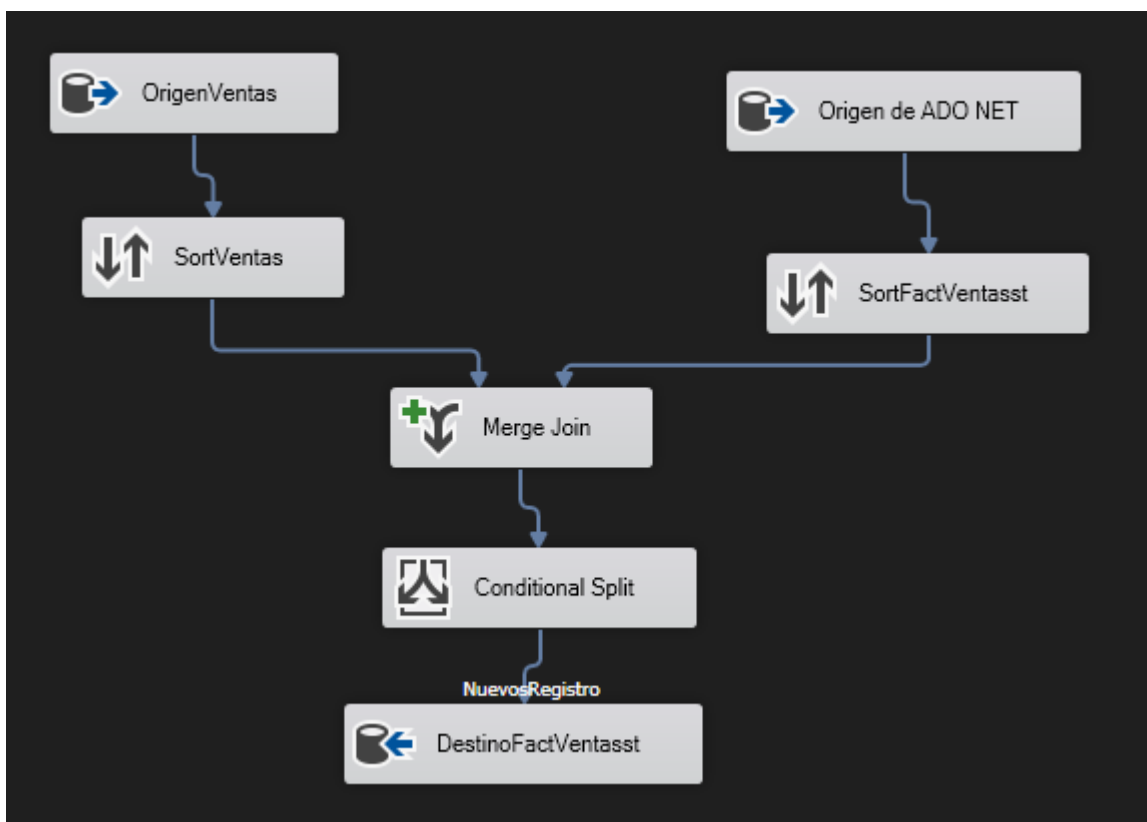


Los datos son llevados de manera correcta a la base de datos:

ID_producto	ID_producto_O	CodigoProducto	nombre	Categoria	precio_venta	proveedor	Desc_Categoria	descripcion_texto
1	1	11679	Sierra de Poda 400MM	2	14.00	HiperGarden Tools	Herramientas	2 para todo tipo de acción
2	2	21636	Pala	2	14.00	HiperGarden Tools	Herramientas	2 para todo tipo de acción
3	3	22225	Rastrillo de Jardín	2	12.00	HiperGarden Tools	Herramientas	2 para todo tipo de acción
4	4	30310	Azadón	2	12.00	HiperGarden Tools	Herramientas	2 para todo tipo de acción
5	14	AR-010	Santolina Chamaecypariss	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
6	12	AR-008	Thymus Citriodora (Tomillo limón)	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
7	11	AR-007	Salvia Mix	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
8	10	AR-006	Petroselinum Hortense (Perejil)	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
9	13	AR-009	Thymus Vulgaris	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
10	8	AR-004	Melissa	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
11	7	AR-003	Mejorana	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
12	6	AR-002	Lavándula Dentata	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
13	5	AR-001	Ajedrea	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
14	9	AR-005	Mentha Sativa	3	1.00	Murcia Seasons	Aromaticas	Plantas aromáticas
15	85	FR-65	Cerezo	4	32.00	Jerte Distribuciones S.L.	Frutales	Árboles pequeños de producción frutal
16	86	FR-66	Cerezo	4	49.00	Jerte Distribuciones S.L.	Frutales	Árboles pequeños de producción frutal
17	87	FR-67	Cerezo	4	70.00	Jerte Distribuciones S.L.	Frutales	Árboles pequeños de producción frutal
18	88	FR-68	Cerezo	4	80.00	Jerte Distribuciones S.L.	Frutales	Árboles pequeños de producción frutal

Tabla fact o hechos:

Para crear la tabla de hechos es necesario tener los datos de las ventas por lo que se agregan a las extracciones las tablas de pedido y detalle pedido luego mediante una consulta se relacionan los datos que serán insertados en la tabla de hechos.



Los datos son llevados de manera correcta a la base de datos:

Results								
	ID_detalle_pedido	ID_detalle_pedidoO	ID_pedido	ID_producto	cantidad	precio_unidad	estado	ID_cliente
1	1	1	1	87	10	70.00	Entregado	5
2	2	2	1	151	40	4.00	Entregado	5
3	3	3	1	165	25	4.00	Entregado	5
4	4	4	1	265	15	19.00	Entregado	5
5	5	5	1	276	23	14.00	Entregado	5
6	6	6	2	57	3	29.00	Entregado	5
7	7	7	2	58	7	8.00	Entregado	5
8	8	8	2	164	50	4.00	Entregado	5
9	9	9	2	165	20	5.00	Entregado	5
10	10	10	2	183	12	6.00	Entregado	5
11	11	11	2	251	67	64.00	Entregado	5
12	12	12	2	271	5	462.00	Entregado	5
13	13	13	3	66	120	9.00	Rechazado	5