



Sri Lanka Institute of Information Technology

**BSc (Hons) in Information Technology Specializing in
Data Science**

Data Warehousing and Business Intelligence IT3021

Assignment 1

2025

Student Name : Madusanka H.P.I

Student Number : IT22259134

Batch : Y3.S1.WE.DS.01

Step 1: Data Set Selection

Chinook Database

Dataset Selected – [Chinook Dataset](#) (click on the text to view the original dataset)

The selected dataset contains digital media store transaction records from the Chinook database, a sample dataset modeled after a digital music store. It includes information on customers, invoices, invoice items, employees, artists, albums, tracks, genres, and media types. The dataset simulates a real-world transactional environment, primarily focused on music purchases, and supports both physical and digital distribution models. Data is organized into multiple interrelated tables such as Customers, Invoices, Tracks, Albums, and Employees, allowing for comprehensive analysis of sales, customer behavior, and business performance. The dataset is well-suited for practicing SQL queries, building ETL workflows, performing data modeling, and developing business intelligence solutions.

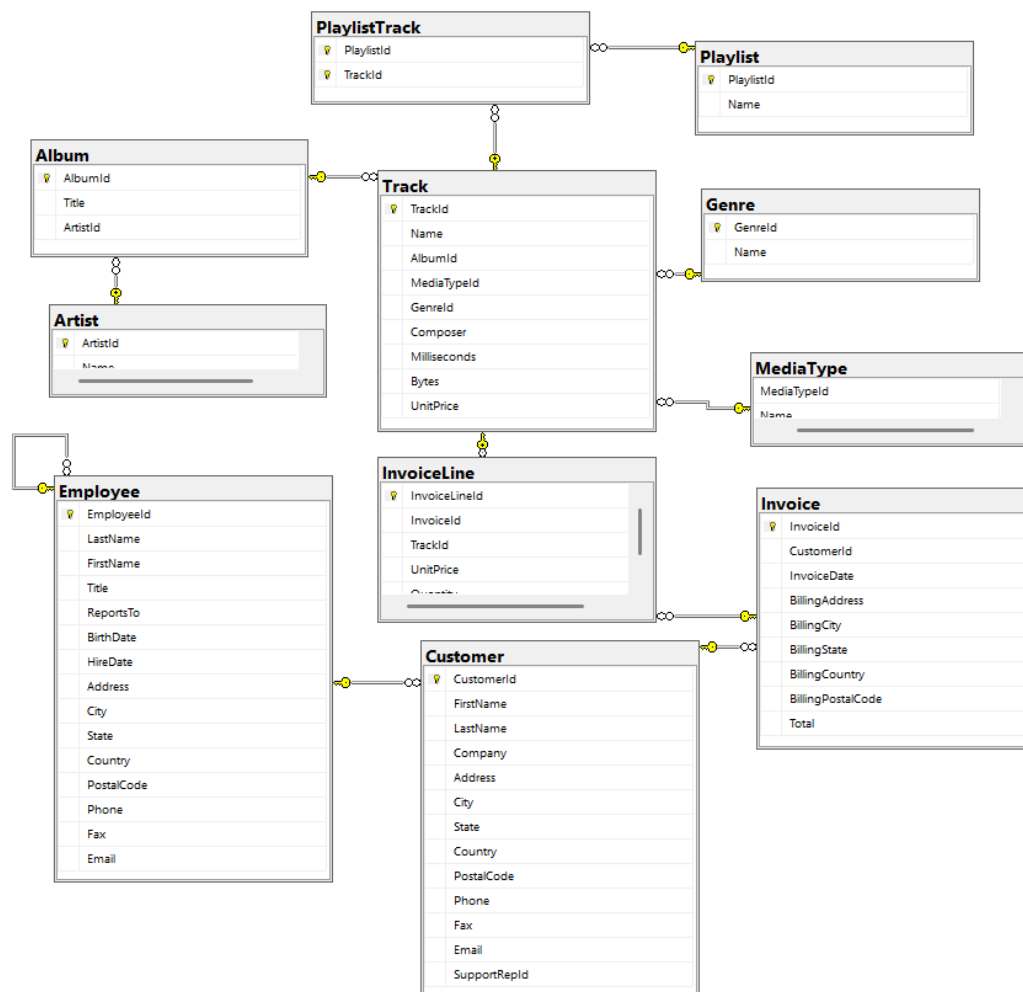


Figure 1: Relationship diagram of the Chinook database

Step 2: Preparation of Data Sources

Database Tables (SQL Server)

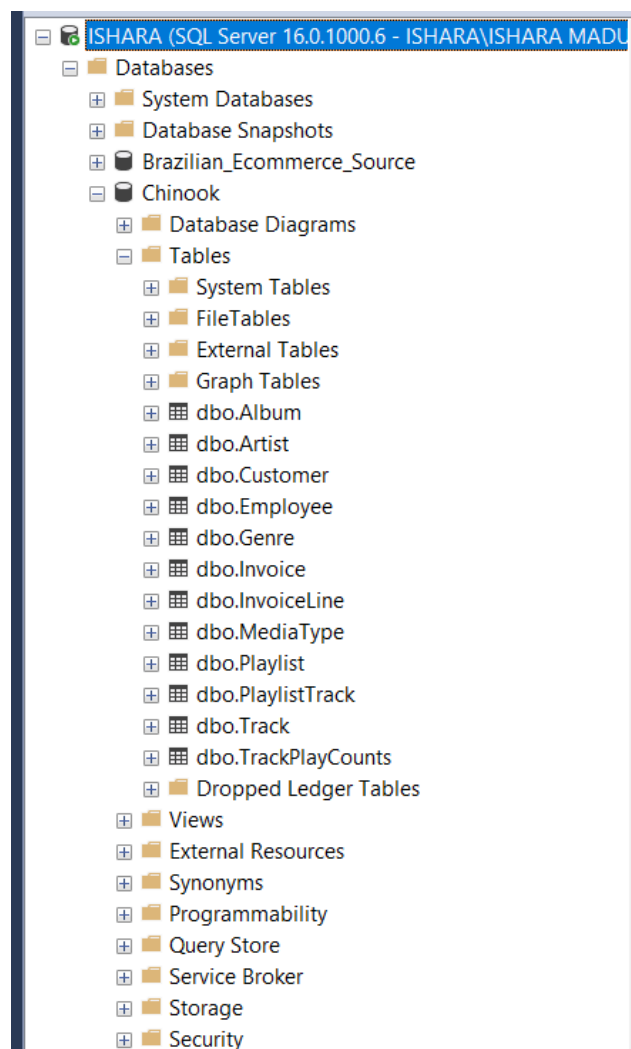
I kept most of the Chinook tables in their original database format including:

Invoice, InvoiceLine, Track, Album, Artist, MediaType, Genre, Customer, Employee

Text Files (.txt)

I exported the following table to text files to serve as a second data source type:

- TrackPlayCounts.txt



Step 3: Solution Architecture

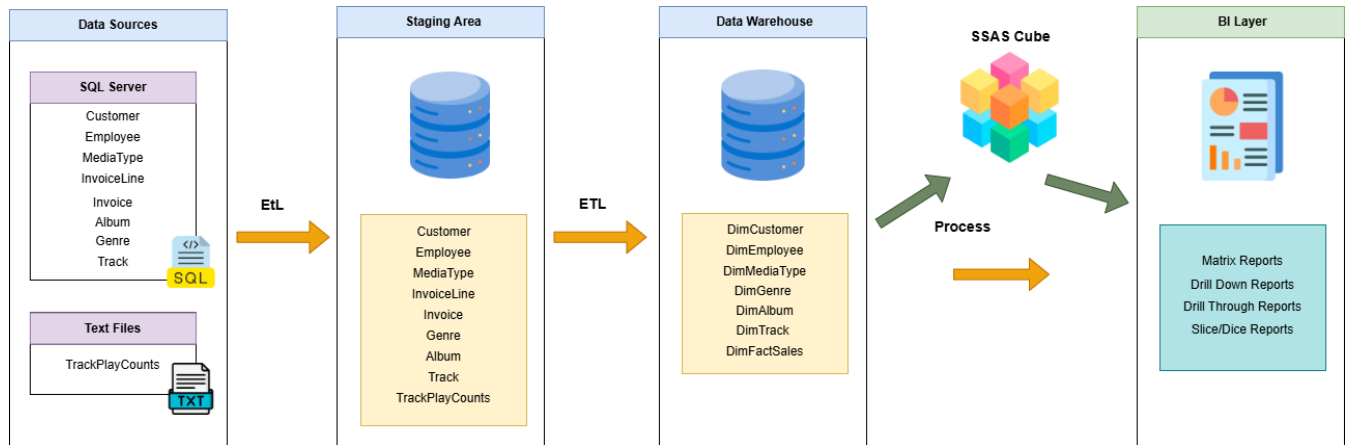


Figure 2: High-level architecture of the Chinook Data Warehouse and BI solution

The architecture consists of the following components:

1. Source Systems:

- **Chinook OLTP Database:** Contains transactional data.
- **Text Files:** Includes additional data sources like TrackPlayCounts.txt

2. Staging Area:

- **Staging Database:** Temporary storage for extracted data.
- **Staging Tables:** Includes tables like StgCustomer, StgInvoice, StgTrack, etc.

3. ETL Process:

- **SQL Server Integration Services (SSIS):** Platform for extracting, transforming, and loading data.
- **ETL Packages:** Includes packages for loading dimensions and facts.

4. Data Warehouse (Dimensional Model):

- **Fact Tables:** FactSales.
- **Dimension Tables:** DimDate, DimCustomer, DimEmployee, etc.

5. **Analytical Processing:**

- **SQL Server Analysis Services (SSAS):** Used for OLAP cube creation.
- **Cubes:** Sales cube for analysis.

6. **Reporting Layer:**

- **Reporting Services:** Used for generating reports.
- **Dashboards:** Includes reports on sales, customer analysis, and track popularity.

Data Flow:

1. Data is extracted and loaded into the staging area.
2. Data is cleaned, transformed, and loaded into dimension tables.
3. Fact data is loaded after dimensions to maintain integrity.
4. Accumulating fact data is updated with transaction completion info.
5. SSAS cubes are processed for querying.
6. Reports are generated from the cubes and data warehouse.

Step 4: Data Warehouse Design & Development

Dimensional Model

I designed a **snowflake schema** for the Chinook data warehouse.



Figure 3: snowflake schema dimensional model for the Chinook data warehouse

Dimension Tables

1. DimDate

- Primary key: DateKey
- Attributes: FullDate, DayOfWeek, DayName, DayOfMonth, DayOfYear, WeekOfYear, MonthName, MonthOfYear, Quarter, QuarterName, Year, IsWeekend

2. DimCustomer (Type-2 SCD)

- Primary key: CustomerSK
- Natural key: AlternateCustomerID
- Attributes: FirstName, LastName, Company, Address, City, State, Country, PostalCode, Phone, Fax, Email, SupportRepID
- SCD fields: StartDate, EndDate, IsCurrent

3. DimEmployee (Type-2 SCD)

- Primary key: EmployeeSK
- Natural key: AlternateEmployeeID
- Attributes: LastName, FirstName, Title, ReportsTo, BirthDate, HireDate, Address, City, State, Country, PostalCode, Phone, Fax, Email
- SCD fields: StartDate, EndDate, IsCurrent

4. DimTrack

- Primary key: TrackSK
- Natural key: AlternateTrackID
- Attributes: Name, AlbumSK, MediaTypeSK, GenreSK, Composer, Milliseconds, Bytes, UnitPrice

5. DimAlbum

- Primary key: AlbumSK
- Natural key: AlternateAlbumID
- Attributes: Title, ArtistSK

6. DimArtist

- Primary key: ArtistSK
- Natural key: AlternateArtistID
- Attributes: Name

7. DimGenre

- Primary key: GenreSK
- Natural key: AlternateGenreID
- Attributes: Name

8. DimMediaType

- Primary key: MediaTypeSK
- Natural key: AlternateMediaTypeID
- Attributes: Name

Fact Table

FactSales

- Primary key: SalesSK
- Foreign keys: CustomerSK, TrackSK, EmployeeSK, DateKey
- Natural keys: InvoiceID, InvoiceLineID
- Measures: UnitPrice, Quantity, ExtendedPrice
- Accumulating fact fields: accm_txn_create_time, accm_txn_complete_time, txn_process_time_hours

Hierarchies

- **Time Hierarchy:** Year → Quarter → Month → Day
- **Geographic Hierarchy:** Country → State → City
- **Music Hierarchy:** Artist → Album → Track

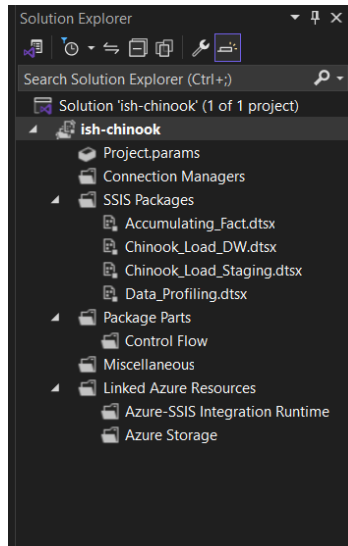
Design Assumptions

1. **Customer and Employee Dimensions:** These are implemented as Type-2 SCDs to track changes in addresses, contact information, and job titles over time.
2. **Track, Album, Artist, Genre, and MediaType Dimensions:** These are implemented as Type-1 SCDs as historical changes are less important for these entities.
3. **Date Dimension:** A complete date dimension was built to support time-based analysis at various levels of granularity.
4. **Multiple Foreign Keys in FactSales:** The fact table includes foreign keys to all relevant dimensions to enable multidimensional analysis.
5. **Surrogate Keys:** Surrogate keys were used for all dimension tables to improve performance and handle SCD scenarios.

Step 5: ETL Development

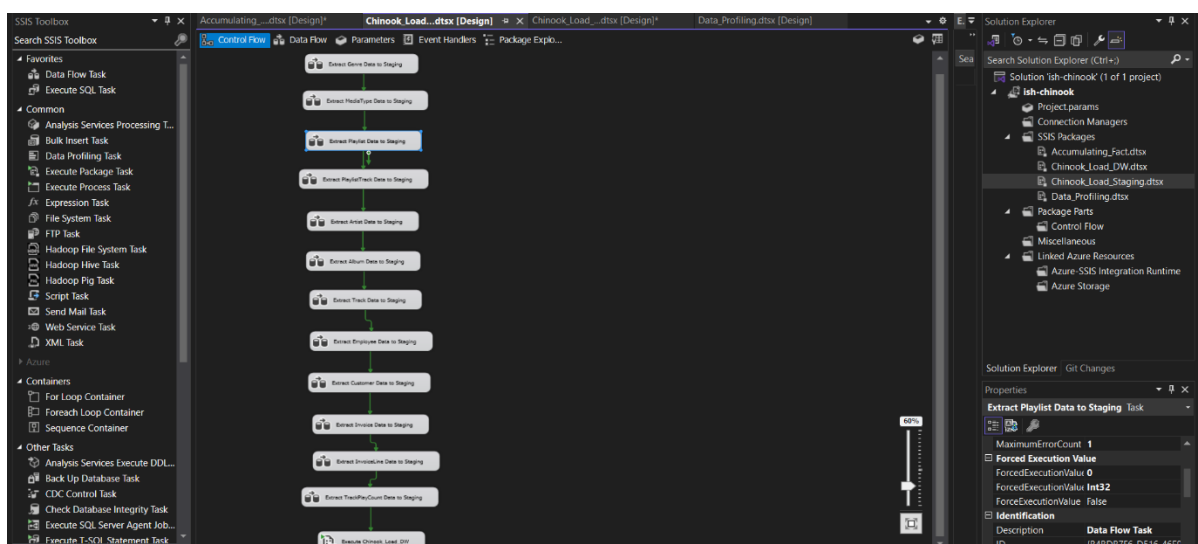
ETL Process Overview

I developed a series of SSIS packages to extract data from multiple source types, transform it according to the data warehouse schema, and load it into the dimensional model:



ETL Packages

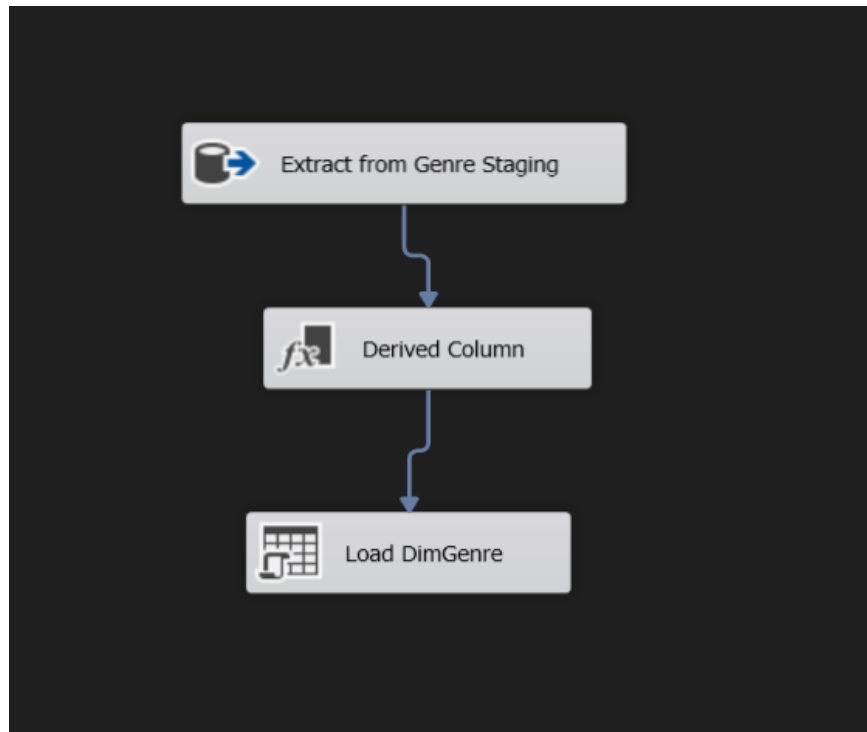
- **Chinook_Load_Staging.dtsx**
 - Extract Data from Source to Staging



▪ Chinook_Load_DW.dtsx

- Loads the Type-1 SCD dimensions (Genre, MediaType, Artist, Album, Track)

1. Genre



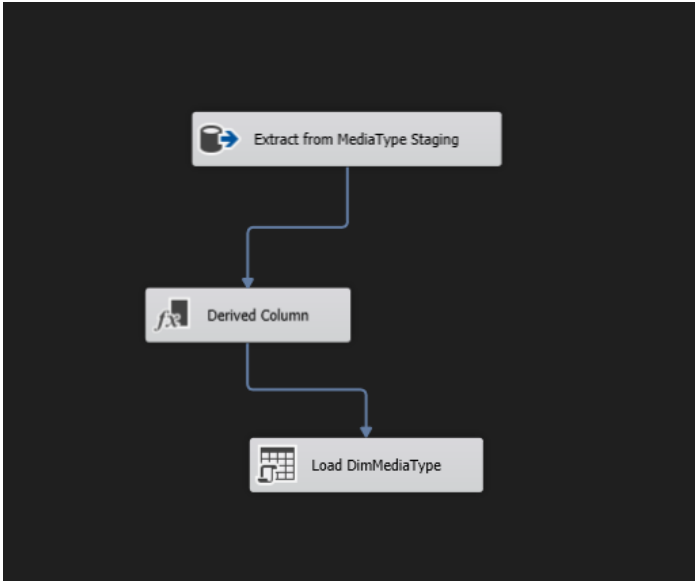
Results Messages

	GenreSK	AlternateGenreID	Name	InsertDate	ModifiedDate
1	1	1	Rock	2025-04-17 17:31:40.110	2025-04-27 20:51:31.260
2	2	2	Jazz	2025-04-17 17:31:40.120	2025-04-27 20:51:31.260
3	3	3	Metal	2025-04-17 17:31:40.120	2025-04-27 20:51:31.260
4	4	4	Alternative & Punk	2025-04-17 17:31:40.123	2025-04-27 20:51:31.260
5	5	5	Rock And Roll	2025-04-17 17:31:40.123	2025-04-27 20:51:31.260
6	6	6	Blues	2025-04-17 17:31:40.123	2025-04-27 20:51:31.260
7	7	7	Latin	2025-04-17 17:31:40.127	2025-04-27 20:51:31.263
8	8	8	Reggae	2025-04-17 17:31:40.127	2025-04-27 20:51:31.263
9	9	9	Pop	2025-04-17 17:31:40.130	2025-04-27 20:51:31.263
10	10	10	Soundtrack	2025-04-17 17:31:40.130	2025-04-27 20:51:31.263
11	11	11	Bossa Nova	2025-04-17 17:31:40.130	2025-04-27 20:51:31.267
12	12	12	Easy Listening	2025-04-17 17:31:40.130	2025-04-27 20:51:31.267
13	13	13	Heavy Metal	2025-04-17 17:31:40.130	2025-04-27 20:51:31.267
14	14	14	R&B/Soul	2025-04-17 17:31:40.133	2025-04-27 20:51:31.267
15	15	15	Electronica/Dance	2025-04-17 17:31:40.133	2025-04-27 20:51:31.267
16	16	16	World	2025-04-17 17:31:40.133	2025-04-27 20:51:31.267
17	17	17	Hip Hop/Rap	2025-04-17 17:31:40.133	2025-04-27 20:51:31.270
18	18	18	Science Fiction	2025-04-17 17:31:40.137	2025-04-27 20:51:31.270
19	19	19	TV Shows	2025-04-17 17:31:40.137	2025-04-27 20:51:31.270
20	20	20	Sci-Fi & Fantasy	2025-04-17 17:31:40.137	2025-04-27 20:51:31.270

Query executed successfully.

ISHARA (16.0 RTM) | ISHARA\ISHARA MADUSANK... | Chinook_DW | 00:00:00 | 25 rows

2. MediaType

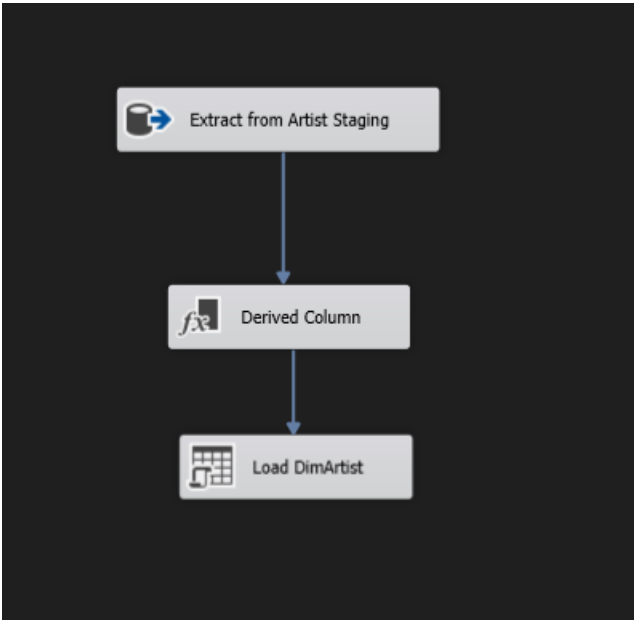


110 %

Results Messages

	MediaTypeSK	AlternateMediaTypeID	Name	InsertDate	ModifiedDate
1	1	1	MPEG audio file	2025-04-17 17:38:34.320	2025-04-27 20:51:31.363
2	2	2	Protected AAC audio file	2025-04-17 17:38:34.330	2025-04-27 20:51:31.367
3	3	3	Protected MPEG-4 video file	2025-04-17 17:38:34.330	2025-04-27 20:51:31.367
4	4	4	Purchased AAC audio file	2025-04-17 17:38:34.333	2025-04-27 20:51:31.367
5	5	5	AAC audio file	2025-04-17 17:38:34.333	2025-04-27 20:51:31.367

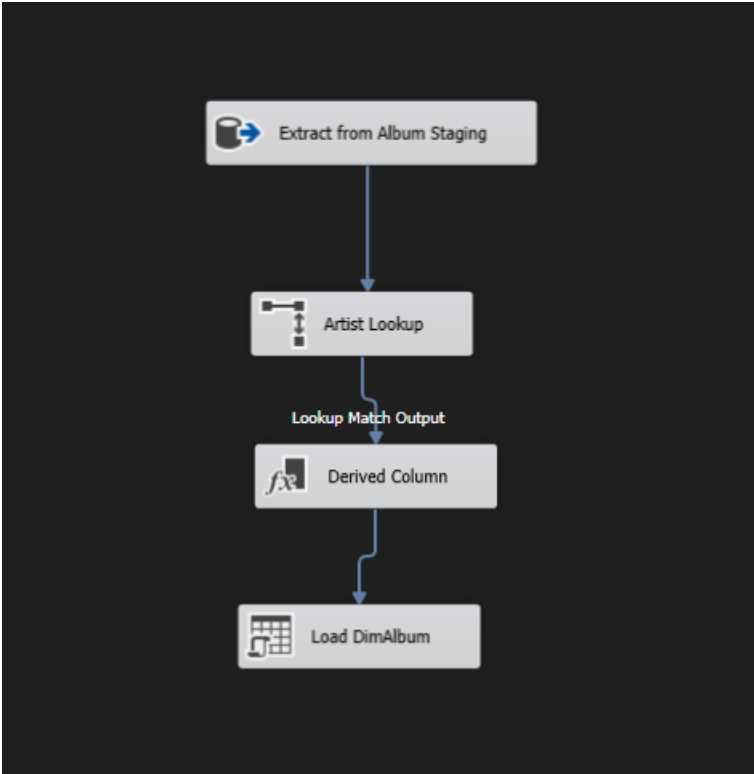
3. Artist



Results Messages					
	ArtistSK	AlternateArtistID	Name	InsertDate	ModifiedDate
1	1	1	AC/DC	2025-04-17 17:41:52.990	2025-04-27 20:51:31.457
2	2	2	Accept	2025-04-17 17:41:53.000	2025-04-27 20:51:31.460
3	3	3	Aerosmith	2025-04-17 17:41:53.000	2025-04-27 20:51:31.460
4	4	4	Alanis Morissette	2025-04-17 17:41:53.003	2025-04-27 20:51:31.460
5	5	5	Alice In Chains	2025-04-17 17:41:53.003	2025-04-27 20:51:31.460
6	6	6	Antônio Carlos Jobim	2025-04-17 17:41:53.003	2025-04-27 20:51:31.460
7	7	7	Apocalyptica	2025-04-17 17:41:53.003	2025-04-27 20:51:31.460
8	8	8	Audioslave	2025-04-17 17:41:53.007	2025-04-27 20:51:31.463
9	9	9	BackBeat	2025-04-17 17:41:53.007	2025-04-27 20:51:31.463
10	10	10	Billy Cobham	2025-04-17 17:41:53.007	2025-04-27 20:51:31.463
11	11	11	Black Label Society	2025-04-17 17:41:53.007	2025-04-27 20:51:31.463
12	12	12	Black Sabbath	2025-04-17 17:41:53.010	2025-04-27 20:51:31.463
13	13	13	Body Count	2025-04-17 17:41:53.010	2025-04-27 20:51:31.467
14	14	14	Bruce Dickinson	2025-04-17 17:41:53.010	2025-04-27 20:51:31.467
15	15	15	Buddy Guy	2025-04-17 17:41:53.010	2025-04-27 20:51:31.467
16	16	16	Caetano Veloso	2025-04-17 17:41:53.013	2025-04-27 20:51:31.467
17	17	17	Chico Buarque	2025-04-17 17:41:53.013	2025-04-27 20:51:31.467
18	18	18	Chico Science & Nação Zumbi	2025-04-17 17:41:53.017	2025-04-27 20:51:31.470
19	19	19	Cidade Negra	2025-04-17 17:41:53.017	2025-04-27 20:51:31.470
20	20	20	Cidade Zélie	2025-04-17 17:41:53.017	2025-04-27 20:51:31.470

Query executed successfully. ISHARA (16.0 RTM) | ISHARA\ISHARA MADUSANK... | Chinook_DW | 00:00:00 | 275 rows

4. Album

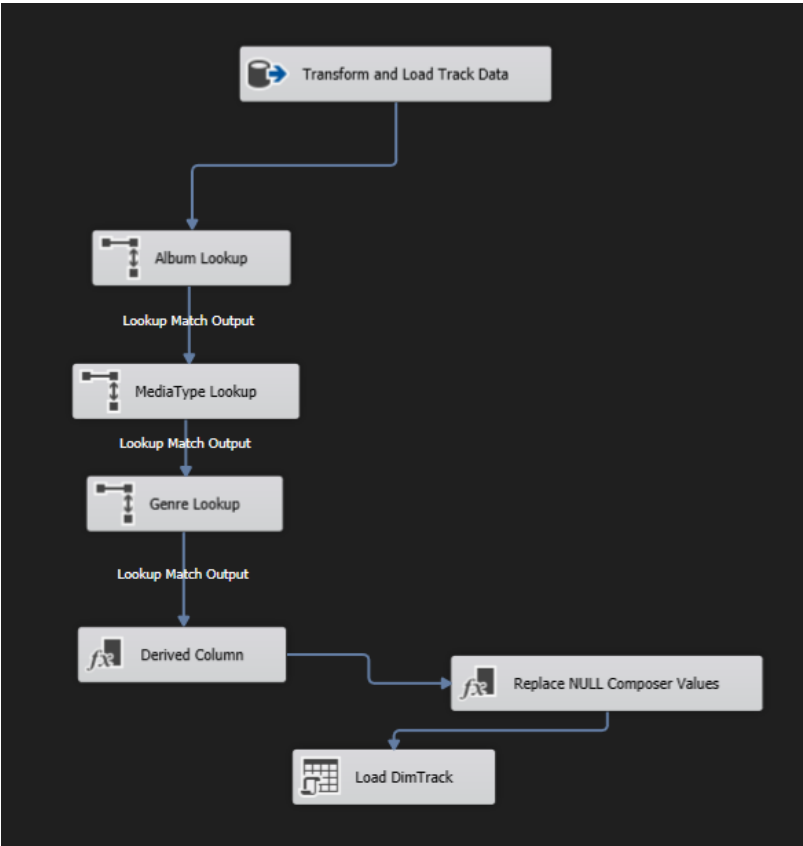


Results		Messages				
	AlbumSK	AlternateAlbumID	Title	ArtistSK	InsertDate	ModifiedDate
1	1	1	For Those About To Rock We Salute You	1	2025-04-17 17:47:46.570	2025-04-27 20:51:32.010
2	2	2	Balls to the Wall	2	2025-04-17 17:47:46.580	2025-04-27 20:51:32.013
3	3	3	Restless and Wild	2	2025-04-17 17:47:46.583	2025-04-27 20:51:32.013
4	4	4	Let There Be Rock	1	2025-04-17 17:47:46.583	2025-04-27 20:51:32.017
5	5	5	Big Ones	3	2025-04-17 17:47:46.587	2025-04-27 20:51:32.017
6	6	6	Jagged Little Pill	4	2025-04-17 17:47:46.587	2025-04-27 20:51:32.020
7	7	7	Facelift	5	2025-04-17 17:47:46.590	2025-04-27 20:51:32.020
8	8	8	Warner 25 Anos	6	2025-04-17 17:47:46.590	2025-04-27 20:51:32.020
9	9	9	Plays Metallica By Four Cellos	7	2025-04-17 17:47:46.590	2025-04-27 20:51:32.023
10	10	10	Audioslave	8	2025-04-17 17:47:46.590	2025-04-27 20:51:32.023
11	11	11	Out Of Exile	8	2025-04-17 17:47:46.593	2025-04-27 20:51:32.027
12	12	12	BackBeat Soundtrack	9	2025-04-17 17:47:46.593	2025-04-27 20:51:32.027
13	13	13	The Best Of Billy Cobham	10	2025-04-17 17:47:46.593	2025-04-27 20:51:32.030
14	14	14	Alcohol Fueled Brewtality Live! [Disc 1]	11	2025-04-17 17:47:46.597	2025-04-27 20:51:32.030
15	15	15	Alcohol Fueled Brewtality Live! [Disc 2]	11	2025-04-17 17:47:46.597	2025-04-27 20:51:32.033
16	16	16	Black Sabbath	12	2025-04-17 17:47:46.597	2025-04-27 20:51:32.033
17	17	17	Black Sabbath Vol. 4 (Remaster)	12	2025-04-17 17:47:46.597	2025-04-27 20:51:32.033
18	18	18	Body Count	13	2025-04-17 17:47:46.600	2025-04-27 20:51:32.037
19	19	19	Chemical Wedding	14	2025-04-17 17:47:46.600	2025-04-27 20:51:32.037
20	20	20	The Best Of Buddy Guy: The Millennium Collection	15	2025-04-17 17:47:46.600	2025-04-27 20:51:32.037

Query executed successfully.

ISHARA (16.0 RTM) | ISHARA\ISHARA MADUSANK... | Chinook_DW | 00:00:00 | 347 rows

5. Track



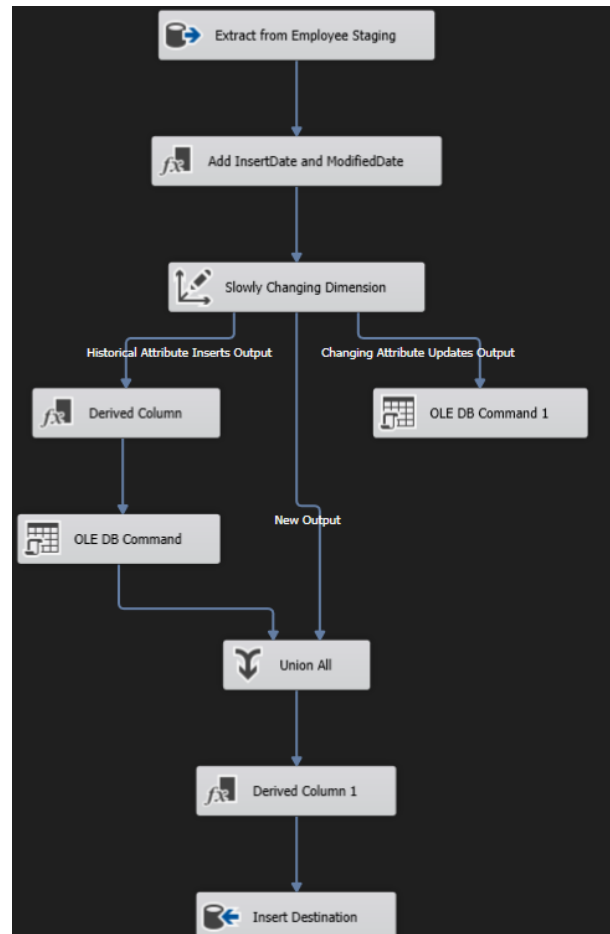
Results Messages												
	TrackSK	AlternateTrackID	Name	AlbumSK	MediaTypeSK	GenreSK	Composer	Milliseconds	Bytes	UnitPrice	InsertDate	ModifiedDate
1	1	1	For Those About To Rock (We Salute You)	1	1	1	Angus Young, Malcolm Young, Brian Johnson	343719	11170334	0.99	2025-04-17 17:58:40.060	2025-04-27
2	2	2	Balls to the Wall	2	2	1	U. Dirkschneider, W. Hoffmann, H. Frank, P. Baltes...	342562	5510424	0.99	2025-04-17 17:58:40.067	2025-04-27
3	3	3	Fast As a Shark	3	2	1	F. Baltes, S. Kaufman, U. Dirkschneider & W. Hoff...	230619	3990994	0.99	2025-04-17 17:58:40.067	2025-04-27
4	4	4	Restless and Wild	3	2	1	F. Baltes, R.A. Smith-Diesel, S. Kaufman, U. Dirks...	252051	4331779	0.99	2025-04-17 17:58:40.070	2025-04-27
5	5	5	Princess of the Dawn	3	2	1	Deafy & R.A. Smith-Diesel	375418	6290521	0.99	2025-04-17 17:58:40.070	2025-04-27
6	6	6	Put The Finger On You	1	1	1	Angus Young, Malcolm Young, Brian Johnson	205662	6713451	0.99	2025-04-17 17:58:40.070	2025-04-27
7	7	7	Let's Get It Up	1	1	1	Angus Young, Malcolm Young, Brian Johnson	233926	7636561	0.99	2025-04-17 17:58:40.070	2025-04-27
8	8	8	Inject The Venom	1	1	1	Angus Young, Malcolm Young, Brian Johnson	210834	6852860	0.99	2025-04-17 17:58:40.070	2025-04-27
9	9	9	Snowballed	1	1	1	Angus Young, Malcolm Young, Brian Johnson	203102	6599424	0.99	2025-04-17 17:58:40.073	2025-04-27
10	10	10	Evil Walks	1	1	1	Angus Young, Malcolm Young, Brian Johnson	263497	8611245	0.99	2025-04-17 17:58:40.073	2025-04-27
11	11	11	C.O.D.	1	1	1	Angus Young, Malcolm Young, Brian Johnson	199836	6566314	0.99	2025-04-17 17:58:40.073	2025-04-27
12	12	12	Breaking The Rules	1	1	1	Angus Young, Malcolm Young, Brian Johnson	263288	8596840	0.99	2025-04-17 17:58:40.077	2025-04-27
13	13	13	Night Of The Long Knives	1	1	1	Angus Young, Malcolm Young, Brian Johnson	205688	6706347	0.99	2025-04-17 17:58:40.077	2025-04-27
14	14	14	Spellbound	1	1	1	Angus Young, Malcolm Young, Brian Johnson	270863	8817038	0.99	2025-04-17 17:58:40.077	2025-04-27
15	15	15	Go Down	4	1	1	AC/DC	331180	10847611	0.99	2025-04-17 17:58:40.077	2025-04-27
16	16	16	Dog Eat Dog	4	1	1	AC/DC	215196	7032162	0.99	2025-04-17 17:58:40.080	2025-04-27
17	17	17	Let There Be Rock	4	1	1	AC/DC	366654	12021261	0.99	2025-04-17 17:58:40.080	2025-04-27
18	18	18	Bad Boy Boogie	4	1	1	AC/DC	267728	8776140	0.99	2025-04-17 17:58:40.080	2025-04-27
19	19	19	Back In Black	4	1	1	AC/DC	336643	10817118	0.99	2025-04-17 17:58:40.080	2025-04-27

Query executed successfully.

ISHARA (16.0 RTM) | ISHARA|ISHARA MADUSANK... | Chinook_DW | 00:00:00 | 1,000 rows

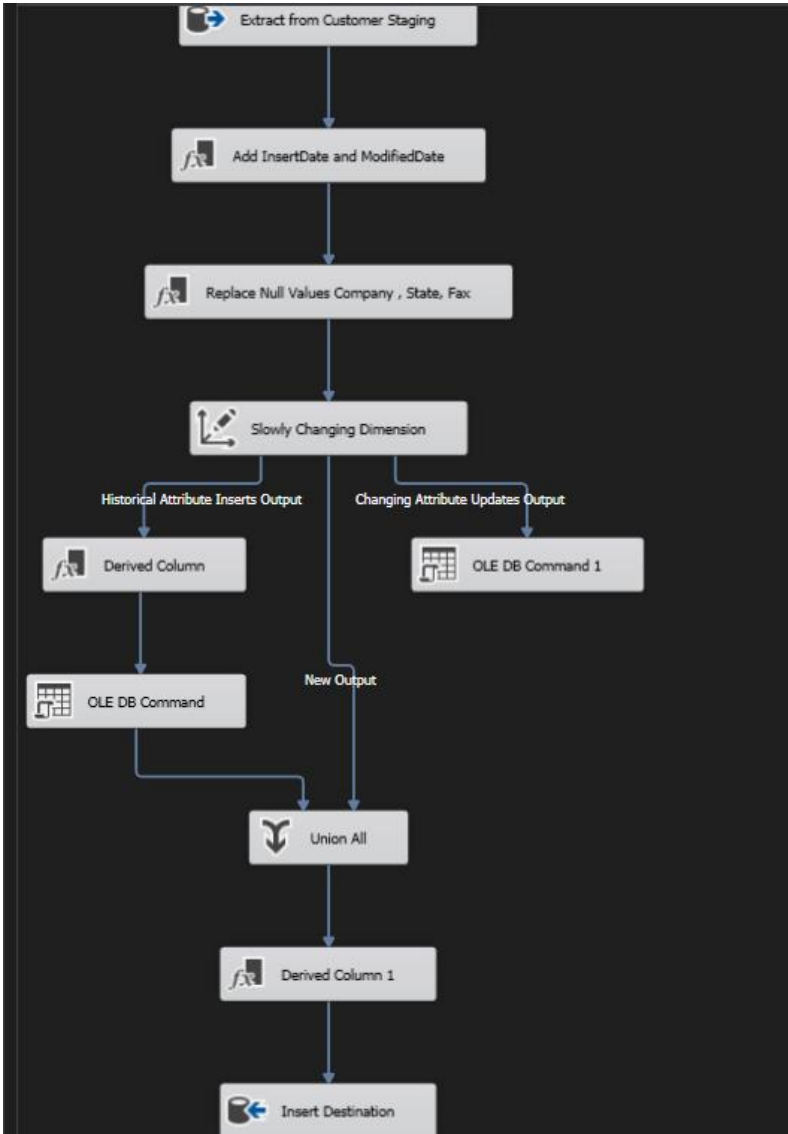
- Loads the Type-2 SCD dimensions (Customer, Employee)

1. Employee



		City	State	Country	PostalCode	Phone	Fax	Email	StartDate	EndDate	IsCurrent	InsertDate	ModifiedDate
1		Matsara	AB	Canada	T2P 5M5	+1 (403) 262-3443	+1 (403) 262-6712	jane@chinookcorp.com	2025-04-17 21:19:17.000	2025-04-27 22:42:29.000	0	2025-04-27 22:10:35.817	2025-04-27 22:10:35.817
2		Matsara2	AB	Canada	T2P 5G3	+1 (403) 263-4423	+1 (403) 263-4289	margaret@chinookcorp.com	2025-04-17 21:19:17.000	2025-04-27 22:45:37.000	0	2025-04-27 22:10:35.817	2025-04-27 22:10:35.817
3		Matsara4	AB	Canada	T3B 1Y7	1 (780) 836-9987	1 (780) 836-9543	steve@chinookcorp.com	2025-04-17 21:19:17.000	2025-04-27 22:49:46.000	0	2025-04-27 22:10:35.817	2025-04-27 22:10:35.817
4		Colombo2	AB	Canada	T5K 2N1	+1 (780) 428-9482	+1 (780) 428-3457	andrew@chinookcorp.com	2025-04-28 09:09:33.000	2025-04-28 09:13:19.000	0	2025-04-28 09:09:38.513	2025-04-28 09:09:38.513
5		Calgary	AB	Canada	T2P 2T3	+1 (403) 262-3443	+1 (403) 262-3322	nancy@chinookcorp.com	2025-04-28 09:09:33.000	2025-04-28 09:14:22.000	0	2025-04-28 09:09:38.513	2025-04-28 09:09:38.513
6		Colombo3	AB	Canada	T2P 5M5	+1 (403) 262-3443	+1 (403) 262-6712	jane@chinookcorp.com	2025-04-28 09:09:33.000	2025-04-28 09:11:31.000	0	2025-04-28 09:09:38.513	2025-04-28 09:09:38.513
7		Calgary	AB	Canada	T2P 5G3	+1 (403) 263-4423	+1 (403) 263-4289	margaret@chinookcorp.com	2025-04-28 09:09:33.000	2025-04-28 09:11:31.000	0	2025-04-28 09:09:38.513	2025-04-28 09:09:38.513
8		Calgary	AB	Canada	T3B 1Y7	1 (780) 836-9987	1 (780) 836-9543	steve@chinookcorp.com	2025-04-28 09:09:33.000	2025-04-28 09:11:31.000	0	2025-04-28 09:09:38.513	2025-04-28 09:09:38.513
9	IW	Calgary	AB	Canada	T3B 0C5	+1 (403) 246-9887	+1 (403) 246-9899	michael@chinookcorp.com	2025-04-28 09:09:33.000	NULL	1	2025-04-28 09:09:38.513	2025-04-28 09:09:38.513
10	rd West	SL	AB	Canada	T1K 5N8	+1 (403) 456-9986	+1 (403) 456-8485	robert@chinookcorp.com	2025-04-28 09:09:33.000	NULL	1	2025-04-28 09:09:38.513	2025-04-28 09:09:38.513
11		MatsaraLk	AB	Canada	T1H 1Y8	+1 (403) 467-3351	+1 (403) 467-8772	laura@chinookcorp.com	2025-04-28 09:09:33.000	NULL	1	2025-04-28 09:09:38.513	2025-04-28 09:09:38.513
12		NewYork	AB	Canada	T2P 5M5	+1 (403) 262-3443	+1 (403) 262-6712	jane@chinookcorp.com	2025-04-28 09:11:31.000	NULL	1	2025-04-28 09:11:36.327	2025-04-28 09:11:36.327
13		Houston	AB	Canada	T2P 5G3	+1 (403) 263-4423	+1 (403) 263-4289	margaret@chinookcorp.com	2025-04-28 09:11:31.000	NULL	1	2025-04-28 09:11:36.327	2025-04-28 09:11:36.327
14		Chicago	AB	Canada	T3B 1Y7	1 (780) 836-9987	1 (780) 836-9543	steve@chinookcorp.com	2025-04-28 09:11:31.000	NULL	1	2025-04-28 09:11:36.327	2025-04-28 09:11:36.327
15		Los Angeles	AB	Canada	T5K 2N1	+1 (780) 428-9482	+1 (780) 428-3457	andrew@chinookcorp.com	2025-04-28 09:13:19.000	NULL	1	2025-04-28 09:13:24.673	2025-04-28 09:13:24.673
16		Boise	AB	Canada	T2P 2T3	+1 (403) 262-3443	+1 (403) 262-3322	nancy@chinookcorp.com	2025-04-28 09:14:22.000	NULL	1	2025-04-28 09:14:27.013	2025-04-28 09:14:27.013

2. Customer

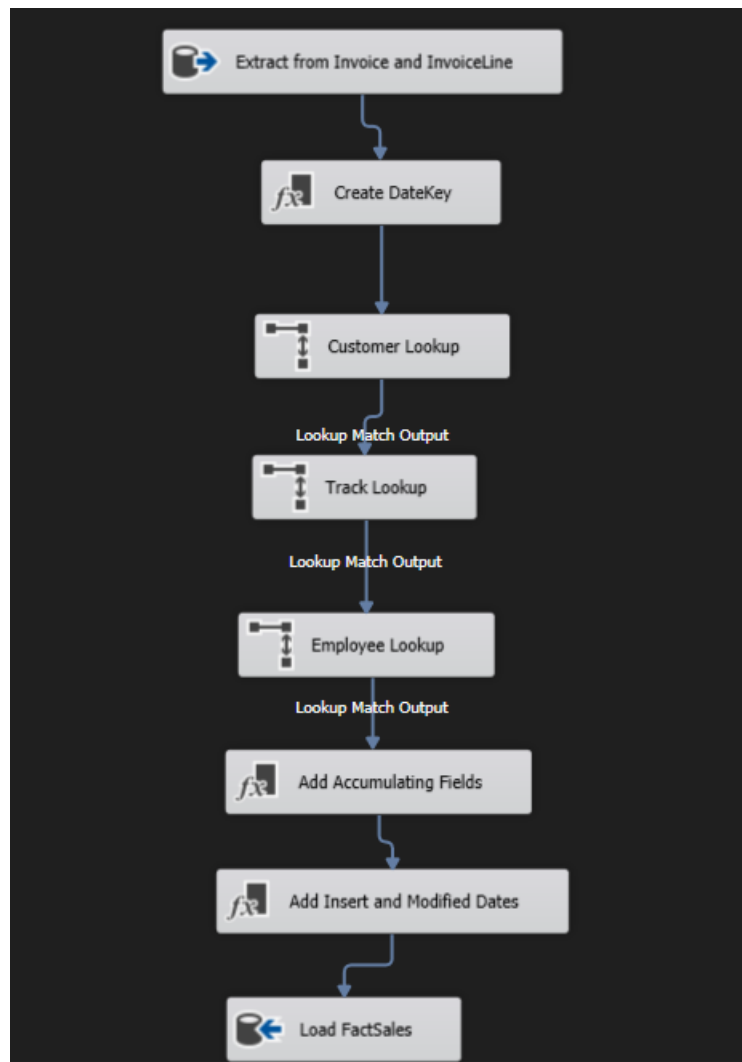


Results Messages													
		City	State	Country	PostalCode	Phone	Fax	Email	SupportRepID	StartDate	EndDate	IsCurrent	
1		o Faria Lima, 2170	Matara	SP	Sri Lanka	12227-000	+55 (12) 3923-5555	+55 (12) 3923-5566	luisg@embraer.com.br	3	2025-04-17 21:10:03.000	2025-04-27 20:46:52.000	0
2		uss-Straße 34	Stuttgart	N	Germany	70174	+49 0711 2842222	N	leonekohler@surfeu.de	5	2025-04-17 21:10:03.000	2025-04-28 08:29:32.000	0
3		langer	Montréal	QC	Canada	H2G 1A7	+1 (514) 721-4711	N	ftremblay@gmail.com	3	2025-04-17 21:10:03.000	2025-04-28 09:50:11.000	0
4		n 14	Oslo	N	Norway	0171	+47 22 44 22 22	N	bjorn.hansen@yahoo.no	4	2025-04-17 21:10:03.000	NULL	1
5		6	Prague	N	Czech Republic	14700	+420 2 4172 5555	+420 2 4172 5555	frantisek@jetbrains.com	4	2025-04-17 21:10:03.000	NULL	1
6		6	Prague	N	Czech Republic	14300	+420 2 4177 0449	N	hholy@gmail.com	5	2025-04-17 21:10:03.000	NULL	1
7		raße 4, 1010 Innere Stadt	Vienne	N	Austria	1010	+43 01 5134505	N	astrid.gruber@apple.at	5	2025-04-17 21:10:03.000	NULL	1
8		33	Brussels	N	Belgium	1000	+32 02 219 03 03	N	daan_peeters@apple.be	4	2025-04-17 21:10:03.000	NULL	1
9		evard 51	Copenhagen	N	Denmark	1720	+453 3331 9991	N	kara.nielsen@jubii.dk	4	2025-04-17 21:10:03.000	NULL	1
10		ão Filho, 155	São Paulo	SP	Brazil	01007-010	+55 (11) 3033-5446	+55 (11) 3033-4564	eduardo@woodstock.com.br	4	2025-04-17 21:10:03.000	NULL	1
11		2022	São Paulo	SP	Brazil	01310-200	+55 (11) 3055-3278	+55 (11) 3055-8131	alero@uol.com.br	5	2025-04-17 21:10:03.000	NULL	1
12		119	Rio de Janeiro	RJ	Brazil	20040-020	+55 (21) 2271-7000	+55 (21) 2271-7070	roberto.almeida@riotur.gov.br	3	2025-04-17 21:10:03.000	NULL	1
13		3	Brasilia	DF	Brazil	71020-677	+55 (61) 3363-5547	+55 (61) 3363-7855	femadaramos4@uol.com.br	4	2025-04-17 21:10:03.000	NULL	1
14		1 NW	Edmonton	AB	Canada	T6G 2C7	+1 (780) 434-4554	+1 (780) 434-5565	mphilips12@shaw.ca	5	2025-04-17 21:10:03.000	NULL	1
15		er Street	Vancouver	BC	Canada	V6C 1G8	+1 (604) 688-2255	+1 (604) 688-8756	jenniferp@rogers.ca	3	2025-04-17 21:10:03.000	NULL	1
16		theatre Parkway	Mountain View	CA	USA	94043-1351	+1 (650) 253-0000	+1 (650) 253-0000	fharris@google.com	4	2025-04-17 21:10:03.000	NULL	1
17		ay	Redmond	WA	USA	98052-8300	+1 (425) 882-8080	+1 (425) 882-8081	jacksmith@microsoft.com	5	2025-04-17 21:10:03.000	NULL	1
18		ay	New York	NY	USA	10012-2612	+1 (212) 221-3546	+1 (212) 221-4679	michelleb@aol.com	3	2025-04-17 21:10:03.000	NULL	1
19													1

Query executed successfully.

ISHARA (16.0 RTM) | ISHARA\ISHARA MADUSANK... | Chinook_DW | 00:00:00 | 62 rows

- Loads the FactSales table after all dimensions are loaded



Results		Messages													
	SalesSK	InvoiceID	InvoiceLineID	CustomerSK	TrackSK	EmployeeSK	DateKey	UnitPrice	Quantity	ExtendedPrice	accm_txn_create_time	accm_txn_complete_time	txn_process_time_hours	InsertDate	M
4	4	2	4	21	8	6	20210102	0.99	1	0.99	2021-01-02 00:00:00.000	2021-01-06 00:00:00.000	96	2025-04-17 21:45:27.790	2
5	5	2	5	21	10	6	20210102	0.99	1	0.99	2021-01-02 00:00:00.000	2021-01-05 00:00:00.000	72	2025-04-17 21:45:27.790	2
6	6	2	6	21	12	6	20210102	0.99	1	0.99	2021-01-02 00:00:00.000	2021-01-08 00:00:00.000	144	2025-04-17 21:45:27.790	2
7	7	3	7	25	16	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-10 00:00:00.000	168	2025-04-17 21:45:27.790	2
8	8	3	8	25	20	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-09 00:00:00.000	144	2025-04-17 21:45:27.790	2
9	9	3	9	25	24	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-08 00:00:00.000	120	2025-04-17 21:45:27.790	2
10	10	3	10	25	28	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-07 00:00:00.000	96	2025-04-17 21:45:27.790	2
11	11	3	11	25	32	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-09 00:00:00.000	144	2025-04-17 21:45:27.790	2
12	12	3	12	25	36	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-06 00:00:00.000	72	2025-04-17 21:45:27.790	2
13	13	4	13	31	42	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-09 00:00:00.000	72	2025-04-17 21:45:27.790	2
14	14	4	14	31	48	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-12 00:00:00.000	144	2025-04-17 21:45:27.790	2
15	15	4	15	31	54	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-10 00:00:00.000	96	2025-04-17 21:45:27.790	2
16	16	4	16	31	60	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-12 00:00:00.000	144	2025-04-17 21:45:27.790	2
17	17	4	17	31	66	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-10 00:00:00.000	96	2025-04-17 21:45:27.790	2
18	18	4	18	31	72	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-11 00:00:00.000	120	2025-04-17 21:45:27.790	2
19	19	4	19	31	78	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-13 00:00:00.000	168	2025-04-17 21:45:27.790	2
20	20	4	20	31	84	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-09 00:00:00.000	72	2025-04-17 21:45:27.790	2
21	21	4	21	31	90	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-10 00:00:00.000	96	2025-04-17 21:45:27.790	2
22	22	4	22	31	96	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-11 00:00:00.000	120	2025-04-17 21:45:27.790	2

Query executed successfully.

ISHARA (16.0 RTM)

ISHARA\ISHARA MADUSANK...

Chinook DW

00:00:00

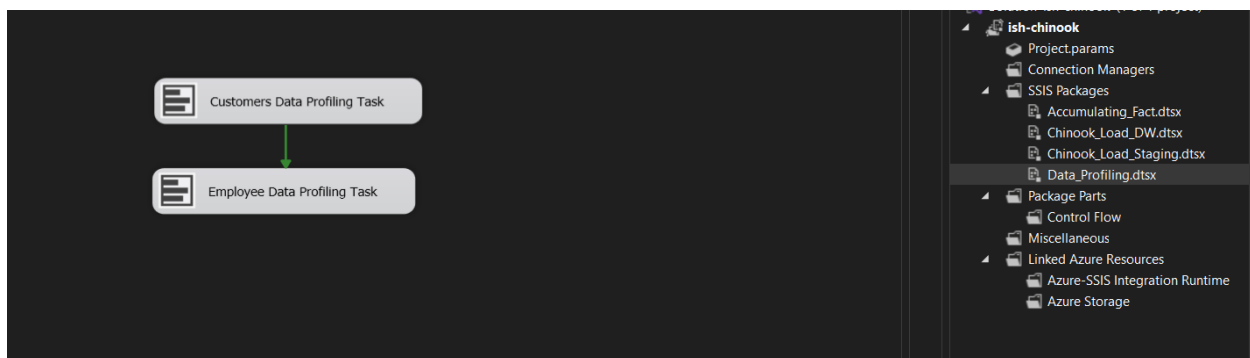
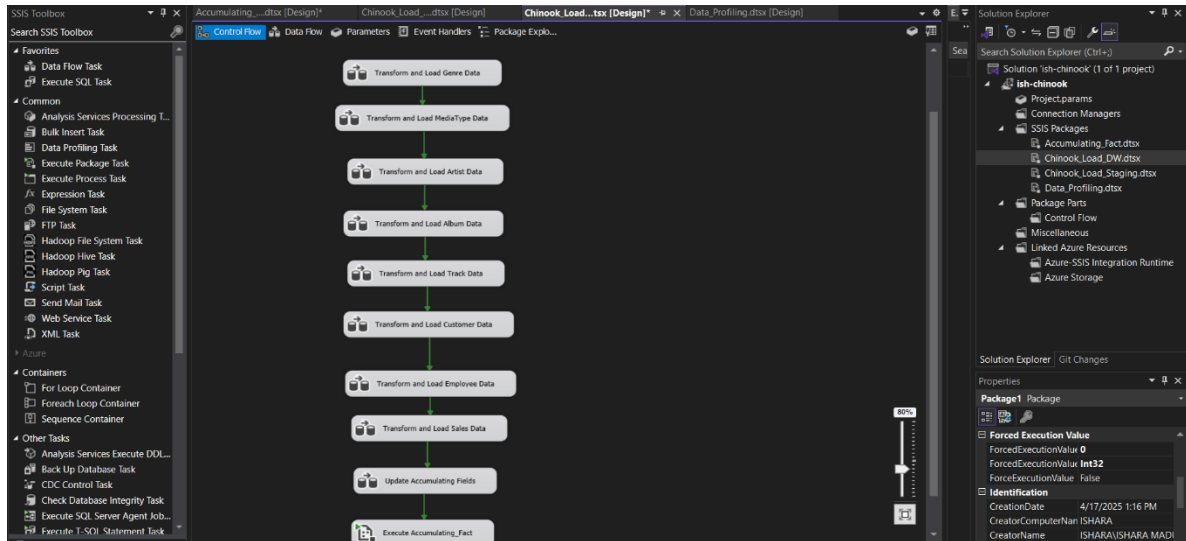
1,000 rows

Query executed successfully.

ISHARA (16.0 RTM) | ISHARA\ISHARA MADUSANK... | Chinook_DW | 00:00:00 | 1,000 rows

▪ Data_Profiling.dtsx

- Check for null or missing values.
- Check data types (integer, string, date, etc.).
- Find minimum, maximum, average values (for numeric data).
- Detect duplicate rows.



▪ Accumulating_Fact.dtsx

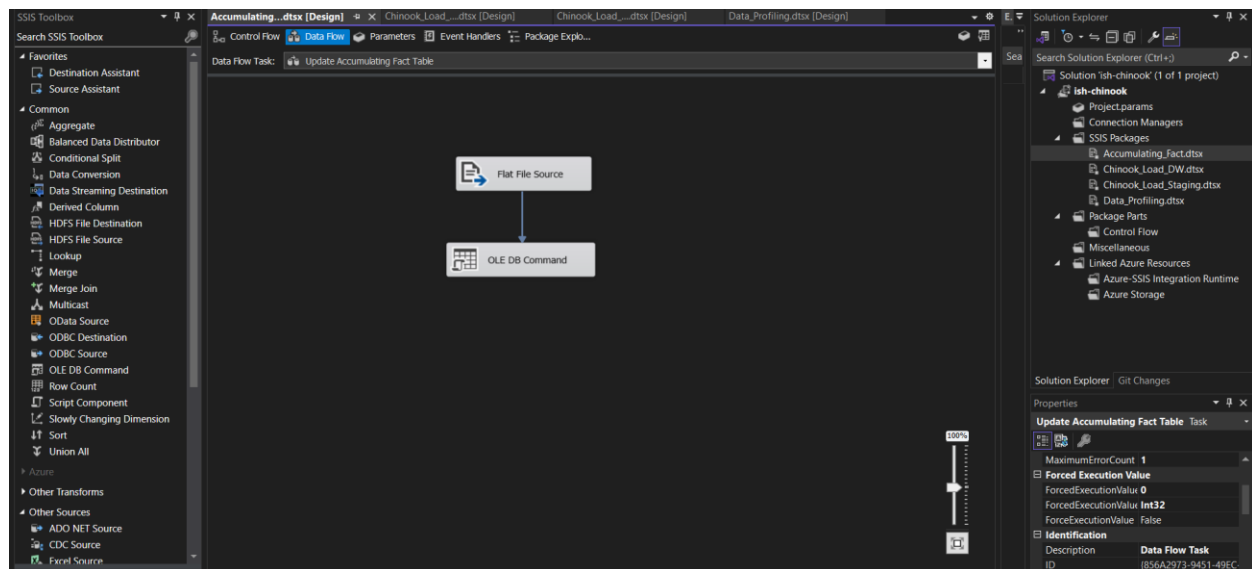
- Updates the accumulating fact fields in FactSales
- Reads from TxnCompletionDates.txt to update completion times

Step 6: ETL Development – Accumulating Fact Tables

Accumulating Fact Implementation

For this requirement, I extended the FactSales table with additional columns to track the lifecycle of transactions:

- **accm_txn_create_time**: Set to the invoice date when records are created
- **accm_txn_complete_time**: Updated later from a separate data source
- **txn_process_time_hours**: Calculated as the difference in hours between create and complete times



Results											
voiceLineID	CustomerSK	TrackSK	EmployeeSK	DateKey	UnitPrice	Quantity	ExtendedPrice	accm_txn_create_time	accm_txn_complete_time	txn_process_time_hours	InsertDate
1	19	2	7	20210101	0.99	1	0.99	2021-01-01 00:00:00.000	2021-01-06 00:00:00.000	120	2025-04-17 21:45:27.790
2	19	4	7	20210101	0.99	1	0.99	2021-01-01 00:00:00.000	2021-01-05 00:00:00.000	96	2025-04-17 21:45:27.790
3	21	6	6	20210102	0.99	1	0.99	2021-01-02 00:00:00.000	2021-01-06 00:00:00.000	96	2025-04-17 21:45:27.790
4	21	8	6	20210102	0.99	1	0.99	2021-01-02 00:00:00.000	2021-01-06 00:00:00.000	96	2025-04-17 21:45:27.790
5	21	10	6	20210102	0.99	1	0.99	2021-01-02 00:00:00.000	2021-01-05 00:00:00.000	72	2025-04-17 21:45:27.790
6	21	12	6	20210102	0.99	1	0.99	2021-01-02 00:00:00.000	2021-01-08 00:00:00.000	144	2025-04-17 21:45:27.790
7	25	16	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-10 00:00:00.000	168	2025-04-17 21:45:27.790
8	25	20	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-09 00:00:00.000	144	2025-04-17 21:45:27.790
9	25	24	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-08 00:00:00.000	120	2025-04-17 21:45:27.790
10	25	28	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-07 00:00:00.000	96	2025-04-17 21:45:27.790
11	25	32	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-09 00:00:00.000	144	2025-04-17 21:45:27.790
12	25	36	6	20210103	0.99	1	0.99	2021-01-03 00:00:00.000	2021-01-06 00:00:00.000	72	2025-04-17 21:45:27.790
13	31	42	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-09 00:00:00.000	72	2025-04-17 21:45:27.790
14	31	48	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-12 00:00:00.000	144	2025-04-17 21:45:27.790
15	31	54	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-10 00:00:00.000	96	2025-04-17 21:45:27.790
16	31	60	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-12 00:00:00.000	144	2025-04-17 21:45:27.790
17	31	66	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-10 00:00:00.000	96	2025-04-17 21:45:27.790
18	31	72	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-11 00:00:00.000	120	2025-04-17 21:45:27.790
19	31	78	7	20210106	0.99	1	0.99	2021-01-06 00:00:00.000	2021-01-12 00:00:00.000	144	2025-04-17 21:45:27.790

Query executed successfully. ISHARA (16.0 RTM) | ISHARA\ISHARA MADUSANK... Chinook_DW | 00:00:00 | 1,000 rows