

Desain dan Implementasi Data Warehouse Penjualan pada Chinook Sample Database

Gusti Ngurah Deva Wirandana Putra^{a1}, Cokorda Pramatha^{b2}

^{a1}Informatics Department, Mathematics and Science Faculty, Udayana University
South Kuta, Badung, Bali, Indonesia

^{b2}Net-Centric Computing Laboratory, Udayana University, Bali, Indonesia

¹wirandanaputra@gmail.com

²cokorda@unud.ac.id

Abstract

The company's decision making is very important for analysis. With the data warehouse can support the process of analysis, design, and business decision making of the company. The company stores operational data that is useful in the business analysis process into a data warehouse. This study will develop the design and implementation of a data warehouse using the Chinook Sample Database as the source. Used Nine-Step Design Methodology to design the data warehouse and through the ETL (Extract, Transformation, Loading) process. The results form a Dashboard that is visualized with Tableau according to a sales fact chart that contains information used to assist the company's business analysis and decisions.

Keywords: Business Intelligence, Data Warehouse, Chinook, Tableau, Nine-Step Design Methodology

1. Pendahuluan

Menurut salah satu tokoh penting di dalam perkembangan dunia teknologi informasi, khususnya dibidang data warehouse, yaitu Willian Harvey Inmon (W.H. Inmon, 1970) atau lebih dikenal dengan Bill Inmon mendefinisikan *data warehouse* sebagai sekumpulan data yang memiliki enam buah sifat atau karakteristik berupa berorientasi subjek (*subject oriented*), terintegrasi (*integrated*), berorientasi pada proses (*process oriented*), *time variant*, dapat diakses dengan mudah (*accessible*), dan bersifat *non-volatile* [1]. Biasanya, organisasi industri menggunakan *data warehouse* untuk melakukan analisa, perencanaan dan pengambilan keputusan bisnis dalam menjalankan proses bisnisnya di suatu perusahaan [2]. Proses bisnis merupakan elemen yang sangat penting dalam membedakan satu perusahaan maupun organisasi satu dengan lainnya, sehingga organisasi tersebut dapat bersaing dengan kompetitornya [3]. Perusahaan akan menyalin data dari penjualan barang ke gudang data atau data yang bersumber dari sistem operasional lainnya, lalu data tersebut akan di analisis dan diberikan *query* yang kompleks sehingga perusahaan dapat memperoleh informasi sesuai yang diinginkan.

Pada tahun 2019, Darmawan Subuh dan Wita Yasman melakukan penelitian mengenai implementasi *data warehouse* dan penerapannya pada Toko Magnifique Clothes dengan menggunakan tools Pentaho untuk melihat kondisi penjualannya. Penelitian tersebut memberikan hasil bahwa dengan dibangunnya *data warehouse* penjualan, maka penyampaian informasi yang terkait dapat dilakukan dengan mudah dan lebih fleksibel. *Data warehouse* dapat digunakan oleh pemilik toko dalam melihat perkembangan keuntungan dan penjualan yang terjadi setiap minggu, bulan dan tahun, sehingga pemilik dapat melakukan analisis terhadap penyampaian informasi yang sudah disajikan dalam bentuk grafik atau dashboard. Informasi yang disajikan dari *data warehouse* penjualan dapat dipergunakan untuk membantu pemilik toko dalam proses evaluasi, perencanaan dan pemasaran barang yang laku terjual, juga membantu toko dalam mencermati tren barang yang diminati konsumen setiap bulan atau tahunnya. *Data warehouse* ini dapat dikembangkan dengan penambahan data-data baru, sehingga *data warehouse* yang ada menjadi tempat penyimpanan dan penyampaian informasi penjualan secara periodik. *Data warehouse* dan *dashboard* aplikasi dapat dibangun berkelanjutan untuk pengembangan ke depannya Toko Magnifique Clothes agar menambah proses lain selain data penjualan, misalnya data pembelian barang ke distributor atau yang lainnya [4].

Pada tahun 2021, I Putu Indie Surya Jayadia, Made Rusdinda Hartania, Wayan Yogi Astira, dan Putu Risky Andreana melakukan penelitian mengenai implementasi *data warehouse* menggunakan Pentaho BI di Hartaning House Homestay. Penelitian ini bertujuan untuk memperoleh informasi,

memberikan prediksi, dan saran, serta kesimpulan yang dapat diberikan berupa penentuan promosi pada tipe kamar yang ada pada Hartaning House. Pada penelitian ini didapatkan hasil bahwa *data warehouse* bermanfaat diimplementasikan dalam segala bidang yang dimiliki salah satunya dalam bidang bisnis pariwisata. *Data warehouse* tidak hanya menjadi gudang data namun dapat digunakan sebagai pembuatan suatu laporan analisis data dengan efisien. Pentaho Business Intelligence dapat digunakan sebagai salah satu solusi yang bermanfaat dalam pembuatan report total penyewaan seluruh jenis kamar pada Hartaning House [5].

Pada penelitian ini, penulis akan melakukan sebuah penelitian pada suatu perusahaan mengenai desain dan implementasi *data warehouse* penjualan. Studi kasus yang digunakan oleh penulis pada penelitian ini yaitu database Chinook yang diperoleh dari situs <https://docs.yugabyte.com/preview/sample-data/chinook/>. Database Chinook merupakan data dari salah satu perusahaan yang bergerak di bidang industri penjualan produk media digital yang terletak di Amerika Serikat. Perusahaan ini menjual 2 tipe produk media yaitu berupa audio dan video. Produk-produk tersebut dimasukkan ke dalam 18 playlist yang diantaranya yaitu music, movies, tv shows, audiobooks, classical, dan lainnya. Perusahaan ini sudah memiliki sebanyak 59 pelanggan yang tersebar di beberapa negara karena produk pada perusahaan ini dapat dibeli secara global dan dari mana saja.

Dengan adanya penerapan *data warehouse* pada database Chinook, maka dapat membantu perusahaan dalam menganalisis data yang sebelumnya masih menggunakan sistem *database* operasional. Selain itu, juga dapat membantu perusahaan dalam mengintegrasikan data yang akan memudahkan perusahaan untuk memahami informasi yang terkandung di dalam data tersebut sehingga perusahaan dapat memperoleh informasi yang bisa digunakan dalam proses pencarian data dan pengambilan keputusan yang terkait dengan penjualan dan pemasok produk dalam suatu perusahaan. Dengan adanya penggunaan *data warehouse* pada perusahaan tersebut, maka diharapkan dapat meringankan kinerja dari sistem *database* operasional yang digunakan dalam operasional perusahaan atau transaksi bisnis perusahaan.

2. Metode Penelitian

2.1. Metode Pengumpulan Data

Pengumpulan data yang digunakan pada penelitian ini didapatkan dengan cara sekunder, dimana tahapannya meliputi studi literatur untuk mendapatkan teori sebagai acuan dalam pembuatan penelitian. Data yang diuji menggunakan *Chinook Sample Database* ini didapatkan dari situs yugabyteDB.

Pada *Chinook Sample Database* terdiri dari beberapa tabel data, antara lain tabel Album, tabel Artist, tabel Customer, tabel Employee, tabel Genre, tabel Invoice, tabel InvoiceLine, tabel MediaType, tabel Playlist, tabel PlaylistTrack, dan tabel Track. Terdapat beberapa data yang ada di dalam masing-masing tabel, antara lain sebagai berikut:

a. Tabel Album

Pada tabel ini terdiri dari 347 baris data yang terdiri dari 3 kolom yang meliputi: AlbumId, Title, dan ArtistId.

AlbumId	Title	ArtistId
1	For Those About To Rock We Salute You	1
2	Balls to the Wall	2
3	Restless and Wild	2
4	Let There Be Rock	1
5	Big Ones	3
6	Jagged Little Pill	4
7	Facelift	5
8	Warner 25 Anos	6
9	Plays Metallica By Four Cellos	7
10	Audioslave	8
11	Out Of Exile	8

Gambar 1. Tabel Album

b. Tabel Artist

Pada tabel ini terdiri dari 275 baris data yang terdiri dari 2 kolom yang meliputi: ArtistId dan Name.

ArtistId	Name
1	AC/DC
2	Accept
3	Aerosmith
4	Alanis Morissette
5	Alice In Chains
6	Antônio Carlos Jobim
7	Apocalyptica
8	Audioslave
9	BackBeat
10	Billy Cobham
11	Black Label Society

Gambar 2. Tabel Artist

c. Tabel Customer

Pada tabel ini terdiri dari 59 baris data yang terdiri dari 13 kolom yang meliputi: CustomerId, FirstName, LastName, Company, Address, City, State, Country, PostalCode, Phone, Fax, Email, dan SupportRepId.

CustomerId	FirstName	LastName	Company	Address	City	State	Country	PostalCode	Phone	Fax	Email	SupportRepId
1	Luis	Gonçalves	Brasileira de Aeronáutica S.A.	Av. Brigadeiro Faria Lima, 2170	São José dos Campos	SP	Brazil	12227-000	+55 (12) 3923-5555	(12) 3923-5566	luisg@embraer.com.br	3
2	Leonie	Köhler	NULL	Theodor-Heuss-Straße 34	Stuttgart	NULL	Germany	70174	+49 0711 2842222	NULL	leoniekohler@surfeu.de	5
3	François	Tremblay	NULL	1498 rue Bélanger	Montréal	QC	Canada	H2G 1A7	+1 (514) 721-4711	NULL	ftremblay@gmail.com	3
4	Bjørn	Hansen	NULL	Ullevålsveien 14	Oslo	NULL	Norway	0171	+47 22 44 22 22	NULL	bjorn.hansen@yahoo.no	4
5	František	Wichterlová	JetBrains s.r.o.	Klanova 9/506	Prague	NULL	Czech Republic	14700	+420 2 4172 5555	2 4172 5555	frantisekw@jetbrains.com	4

Gambar 3. Tabel Customer

d. Tabel Employee

Pada tabel ini terdiri dari 8 baris data yang terdiri dari 15 kolom yang meliputi: ProductID, EmployeeId, LastName, FirstName, Title, ReportsTo, BirthDate, HireDate, Address, City, State, Country, PostalCode, Phone, Fax, Email

EmployeeId	LastName	FirstName	Title	ReportsTo	BirthDate	HireDate	Address	City	State	Country	PostalCode	Phone	Fax	Email
1	Adams	Andrew	General Manager	NULL	1962-02-18 00:00:00	2002-08-14 00:00:00	11120 Jasper Ave NW	Edmonton	AB	Canada	T5K 2N1	+1 (780) 428-9482	+1 (780) 428-3457	andrew@chinookcorp.com
2	Edwards	Nancy	Sales Manager	1	1958-12-08 00:00:00	2002-05-01 00:00:00	825 8 Ave SW	Calgary	AB	Canada	T2P 2T3	+1 (403) 262-3443	+1 (403) 262-3322	nancy@chinookcorp.com
3	Peacock	Jane	Sales Support Agent	2	1973-08-29 00:00:00	2002-04-01 00:00:00	1111 6 Ave SW	Calgary	AB	Canada	T2P 5M5	+1 (403) 262-3443	+1 (403) 262-6712	jane@chinookcorp.com
4	Park	Margaret	Sales Support Agent	2	1947-09-19 00:00:00	2003-05-03 00:00:00	683 10 Street SW	Calgary	AB	Canada	T2P 5G3	+1 (403) 263-4423	+1 (403) 263-4289	margaret@chinookcorp.com
5	Johnson	Steve	Sales Support Agent	2	1965-03-03 00:00:00	2003-10-17 00:00:00	7727B 41 Ave	Calgary	AB	Canada	T3B 1Y7	1 (780) 836-9987	1 (780) 836-9543	steve@chinookcorp.com

Gambar 4. Tabel Employee

e. Tabel Genre

Pada tabel ini terdiri dari 25 baris data yang terdiri dari 2 kolom yang meliputi: GenreId dan Name.

GenreId	Name
1	Rock
2	Jazz
3	Metal
4	Alternative & Punk
5	Rock And Roll
6	Blues
7	Latin
8	Reggae
9	Pop
10	Soundtrack

Gambar 5. Tabel Genre

f. Tabel Invoice

Pada tabel ini terdiri dari 412 baris data yang terdiri dari 9 kolom yang meliputi: InvoiceId, CustomerId, InvoiceDate, BillingAddress, BillingCity, BillingState, BillingCountry, BillingPostalCode, dan Total.

InvoiceId	CustomerId	InvoiceDate	BillingAddress	BillingCity	BillingState	BillingCountry	BillingPostalCode	Total
412	58	2013-12-22 00:00:00	12,Community Centre	Delhi	NULL	India	110017	1.99
411	44	2013-12-14 00:00:00	Porthaninkatu 9	Helsinki	NULL	Finland	00530	13.86
410	35	2013-12-09 00:00:00	Rua dos Campeões Europeus de Viena, 4350	Porto	NULL	Portugal	NULL	8.91
409	29	2013-12-06 00:00:00	796 Dundas Street West	Toronto	ON	Canada	M6J 1V1	5.94
408	25	2013-12-05 00:00:00	319 N. Frances Street	Madison	WI	USA	53703	3.96
406	21	2013-12-04 00:00:00	801 W 4th Street	Reno	NV	USA	89503	1.98
407	23	2013-12-04 00:00:00	69 Salem Street	Boston	MA	USA	2113	1.98
405	20	2013-11-21 00:00:00	541 Del Medio Avenue	Mountain View	CA	USA	94040-111	0.99
404	6	2013-11-13 00:00:00	Rilská 3174/6	Prague	NULL	Czech Republic	14300	25.86
403	56	2013-11-08 00:00:00	307 Macacha Güemes	Buenos Aires	NULL	Argentina	1106	8.91
402	50	2013-11-05 00:00:00	C/ San Bernardo 85	Madrid	NULL	Spain	28015	5.94
401	46	2013-11-04 00:00:00	3 Chatham Street	Dublin	Dublin	Ireland	NULL	3.96

Gambar 6. Tabel Invoice

g. Tabel InvoiceLine

Pada tabel ini terdiri dari 2.240 baris data yang terdiri dari 5 kolom yang meliputi: InvoiceLineId, InvoiceId, TrackId, UnitPrice, dan Quantity.

InvoiceLineId	InvoiceId	TrackId	UnitPrice	Quantity
1	1	2	0.99	1
2	1	4	0.99	1
3	2	6	0.99	1
4	2	8	0.99	1
5	2	10	0.99	1
6	2	12	0.99	1
7	3	16	0.99	1
8	3	20	0.99	1
9	3	24	0.99	1
10	3	28	0.99	1

Gambar 7. Tabel InvoiceLine

h. Tabel MediaType

Pada tabel ini terdiri dari 5 baris data yang terdiri dari 2 kolom yang meliputi: MediaTypeId dan Name.

MediaTypeId	Name
1	MPEG audio file
2	Protected AAC audio file
3	Protected MPEG-4 video file
4	Purchased AAC audio file
5	AAC audio file

Gambar 8. Tabel MediaType

i. Tabel Playlist

Pada tabel ini terdiri dari 18 baris data yang terdiri dari 2 kolom yang meliputi: PlaylistId dan Name.

PlaylistId	Name
1	Music
2	Movies
3	TV Shows
4	Audiobooks
5	90's Music
6	Audiobooks
7	Movies
8	Music
9	Music Videos
10	TV Shows

Gambar 9. Tabel Playlist

j. Tabel PlaylistTrack

Pada tabel ini terdiri dari 8.715 baris data yang terdiri dari 2 kolom yang meliputi: PlaylistId dan TrackId.

PlaylistId	TrackId
1	1
1	2
1	3
1	4
1	5
1	6
1	7
1	8
1	9
1	10

Gambar 10. Tabel PlaylistTrack

k. Tabel Track

Pada tabel ini terdiri dari 3.503 baris data yang terdiri dari 9 kolom yang meliputi: TrackId, Name, AlbumId, MediaTypeId, GenreId, Composer, Milliseconds, Bytes, dan UnitPrice.

TrackId	Name	AlbumId	MediaTypeId	GenreId	Composer	Milliseconds	Bytes	UnitPrice
1	For Those About To Rock (We Salute You)	1	1	1	Angus Young, Malcolm Young, Brian Johnson	343719	11170334	0.99
2	Balls to the Wall	2	2	1	NULL	342562	5510424	0.99
3	Fast As a Shark	3	2	1	F. Baltes, S. Kaufman, U. Dirksneider & W. Hoffma...	230619	3990994	0.99
4	Restless and Wild	3	2	1	F. Baltes, R.A. Smith-Diesel, S. Kaufman, U. Dirks...	252051	4331779	0.99
5	Princess of the Dawn	3	2	1	Deafy & R.A. Smith-Diesel	375418	6290521	0.99
6	Put The Finger On You	1	1	1	Angus Young, Malcolm Young, Brian Johnson	205662	6713451	0.99
7	Let's Get It Up	1	1	1	Angus Young, Malcolm Young, Brian Johnson	233926	7636561	0.99
8	Inject The Venom	1	1	1	Angus Young, Malcolm Young, Brian Johnson	210834	6852860	0.99
9	Snowballed	1	1	1	Angus Young, Malcolm Young, Brian Johnson	203102	6599424	0.99
10	Evil Walks	1	1	1	Angus Young, Malcolm Young, Brian Johnson	263497	8611245	0.99

Gambar 11. Tabel Track

2.2. OLAP (On Line Analytical Processing)

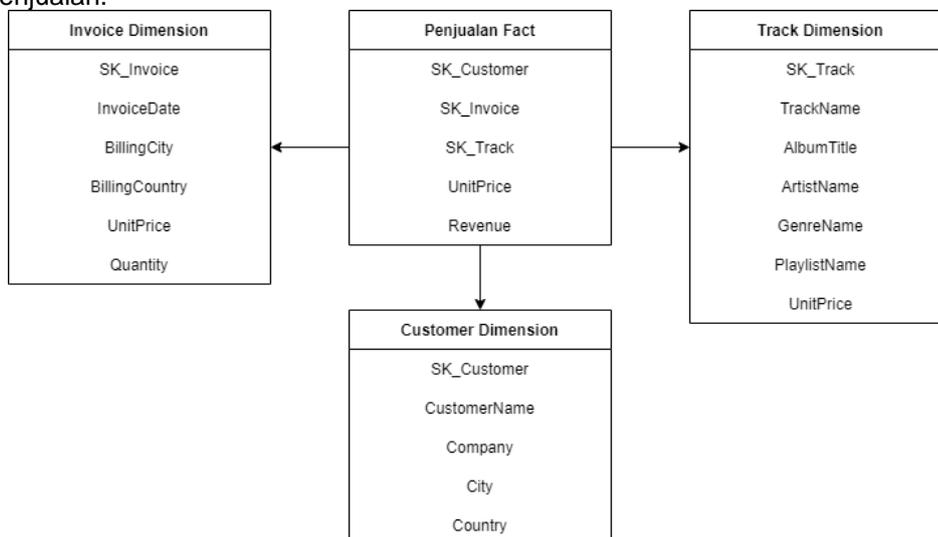
OLAP merupakan suatu proses komputer yang dapat memungkinkan para penggunanya dengan mudah dan selektif dalam memilih dan melihat data dari sudut pandang yang berbeda-beda. Data pada OLAP disimpan dalam basis data multidimensi. Pada basis data multidimensi terdiri dari banyak dimensi yang dapat dipisahkan oleh OLAP menjadi beberapa sub atribut. Berbeda dengan basis data relasional yang hanya terdiri dari dua dimensi [1]. Data multidimensi memiliki tiga atribut untuk bisa dikelola dalam OLAP diantaranya yaitu [6]:

- Dimensi (*Dimension*) adalah suatu atribut yang di tinjau.
- Pengukur (*Measurement*) adalah besaran yang dapat diukur mengacu pada irisan antara dimensi yang di tinjau.
- Hasil Pengukuran/Kalkulasi (*Calculation*) adalah nilai dari measurement

2.3. Perancangan Data Warehouse

Perancangan *data warehouse* yang digunakan dalam penelitian ini menggunakan metode *Nine-Step Design Methodology* yang di dalamnya terdiri dari beberapa langkah, antara lain:

1. Analisis Kebutuhan
Analisis kebutuhan informasi pada perancangan *data warehouse* untuk prediksi penjualan produk menggunakan data dari *Chinook Sample Database* yang di dalamnya berisi informasi pelanggan, data transaksi pelanggan, track, album, artis, genre, dan playlist.
2. Memilih Proses (*Choosing the Process*)
Pada tahap ini bertujuan untuk menentukan proses bisnis dari perusahaan yang ingin membangun *data warehouse*. Berdasarkan informasi yang diperoleh dari analisis kebutuhan, proses bisnis yang terjadi pada penelitian ini adalah penjualan produk.
3. Memilih *Grain* (*Choosing the Grain*)
Pada tahap ini bertujuan untuk memilih *grain* yang digunakan sebagai dasar sebelum membuat tabel fakta (*fact table*). Berdasarkan proses bisnis yang ditentukan, *grain* yang dipilih pada penelitian ini adalah *unit price* atau harga produk yang dibeli *customer*.
4. Mengidentifikasi dan Penyesuaian Dimensi (*Identifying and Conforming the Dimensions*)
Pada tahap ini bertujuan untuk mengidentifikasi dan memilih dimensi yang berhubungan dengan tabel fakta. Dari hasil identifikasi, maka tabel dimensi yang berhubungan dengan tabel fakta meliputi:
 - a. Dimensi Customer
Dimensi Customer merupakan pembelian produk yang dilakukan oleh customer. *Grain* yang berhubungan dengan dimensi ini adalah customer yang melakukan pembelian produk.
 - b. Dimensi Track
Dimensi Track merupakan produk yang tersedia di *Chinook Sample Database*. *Grain* yang berhubungan dengan dimensi ini adalah jumlah produk/track yang dibeli,
 - a. Dimensi Invoice
Dimensi Invoice merupakan data transaksi tersimpan dalam database. *Grain* yang berhubungan dengan dimensi ini adalah jumlah transaksi yang terjadi.
5. Memilih Fakta (*Choosing the Fact*)
Tahap ini bertujuan untuk memilih tabel fakta (*fact table*) berdasarkan *grain* yang sudah ditentukan sebelumnya. Setelah mendapatkan *grain*, maka tabel fakta yang dipilih adalah tabel fakta penjualan.



Gambar 12. Model Tabel Fakta

Pada gambar 12 menggunakan model tabel fakta yaitu model skema bintang (*Star Schema*). Skema bintang ini memiliki hubungan tabel fakta dan tabel dimensi menyerupai bintang. Skema ini memiliki keuntungan yaitu dapat meningkatkan kinerja *data warehouse*, pemrosesan *query* yang lebih efisien, dan waktu respon yang cepat.

6. Menyimpan pre-Calculation dalam Tabel Fakta (*Storing Pre-Calculation in the Fact Table*)
Agregasi pada tabel fakta penjualan adalah total jumlah produk yang dibeli berdasarkan waktu (hari, bulan dan tahun). *Pre-Calculation* yang ada dalam tabel fakta yaitu *revenue* dari penjualan.
7. Melengkapi Tabel Dimensi (*Rounding Out the Dimension Tables*)
Pada tahap ini bertujuan untuk melengkapi atribut yang ada dalam masing-masing tabel dimensi. Pada tabel dimensi yang sebelumnya telah ditentukan, masing-masing tabel dimensi memiliki atribut sebagai berikut:

Table 1. Dimensi Track

Field	Size	Keterangan
SK_Track	integer(11)	Nomor identitas track
TrackName	varchar(200)	Nama track
AlbumTitle	varchar(160)	Judul album pada track
ArtistName	varchar(120)	Nama artis pada track
GenreName	varchar(120)	Nama genre pada track
PlaylistName	varchar(120)	Nama playlist pada track
UnitPrice	decimal(10,2)	Harga satuan track

Table 2. Dimensi Customer

Field	Size	Keterangan
SK_Customer	integer(11)	Nomor identitas customer
CustomerName	varchar(60)	Nama customer
Company	varchar(80)	Nama perusahaan customer
City	varchar(40)	Kota tinggal customer
Country	varchar(40)	Negara tinggal customer

Table 3. Dimensi Invoice

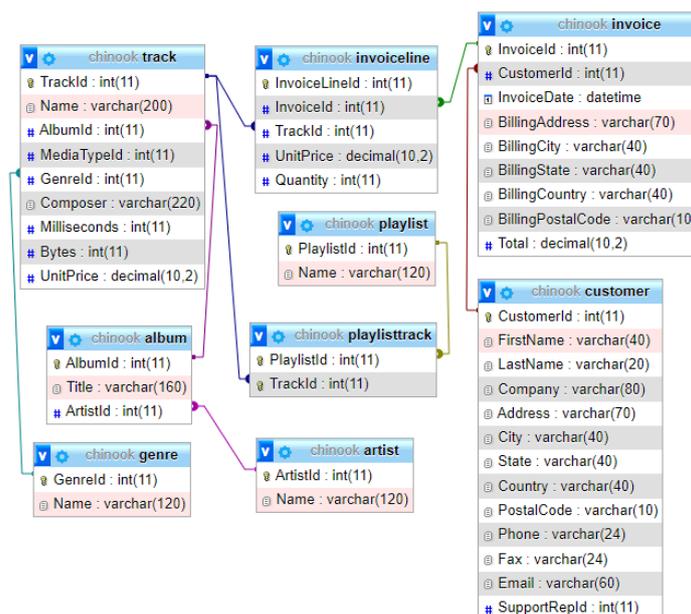
Field	Size	Keterangan
SK_Invoice	integer(11)	Nomor identitas invoice
InvoiceDate	datetime	Tanggal track dibeli customer
BillingCity	varchar(40)	Kota tempat pembelian
BillingCountry	varchar(40)	Negara tempat pembelian
UnitPrice	decimal(10,2)	Harga satuan track
Quantity	decimal(10,2)	Jumlah produk yang dibeli

8. Pemilihan Durasi Database (*Choosing the Duration of Database*)
Tahap ini bertujuan untuk memilih durasi waktu yang digunakan dalam perancangan *data warehouse* yang akan dibangun. Dalam perancangan ini, durasi waktu yang digunakan selama lima tahun terakhir yaitu dari tahun 2009-2013.
9. Melacak Perubahan Dari Dimensi
Atribut yang terdapat pada tabel dimensi bisa saja mengalami perubahan yang dinamis. Perubahan tersebut terjadi karena adanya proses ETL pada database penelitian.

3. Hasil dan Pembahasan

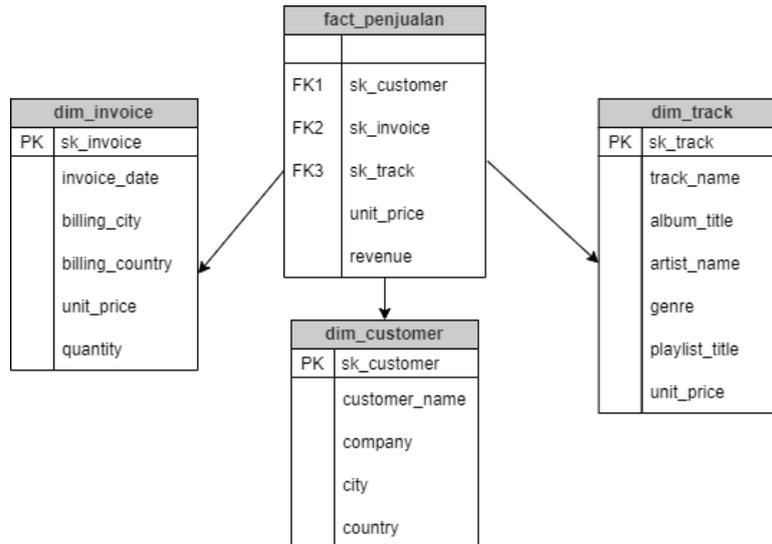
3.1. Skema Database Chinook

Dalam penelitian ini, penulis menggunakan sembilan tabel data dari database Chinook. Tabel-tabel tersebut diantaranya yaitu: tabel invoice, invoiceLine, track, playlist, album, playlisttrack, genre, artist, dan customer. Tabel-tabel tersebut kemudian direlasikan sehingga membentuk sebuah skema *database*. Selanjutnya skema *database* ini akan dianalisis untuk merancang skema *data warehouse* penjualan.



Gambar 13. Skema Database Chinook

3.2. Skema Data Warehouse



Gambar 14. Skema Data Warehouse

Berdasarkan analisis dari skema *database* Chinook, penulis merancang sebuah skema untuk *data warehouse* penjualan. Terdapat satu *fact tabel* di dalam skema tersebut dengan nama *fact penjualan* yang merupakan representasi dari jumlah transaksi penjualan yang terdapat pada *database* Chinook yang berdasarkan tabel dimensi customer, tabel dimensi track dan tabel dimensi invoice. Selain itu, di dalam tabel dimensi juga menyimpan beberapa atribut yang terkait. *Fact penjualan* akan menyimpan agregasi/kalkulasi dari jumlah barang yang terjual pada perusahaan dan pendapatan yang dihasilkan dari penjualan produk perusahaan. Sehingga dari hasil kalkulasi tersebut, dapat diperoleh informasi yang dapat membantu perusahaan.

3.3. Proses ETL (Extract, Transformation, Loading)

Proses ETL ini merupakan proses integrasi data dari berbagai sumber data untuk menghasilkan sudut pandang tunggal terhadap semua data yang diintegrasikan tersebut. Pada aktivitas ini, terdapat tahapan proses yang harus dilakukan, yaitu

1. *Extraction* merupakan pemilahan dan pengambilan data dari satu atau lebih sumber data.
2. *Cleansing* merupakan pembersihan data untuk meyakinkan validitas, kualitas, dan konsistensi antar data serta penghilangan duplikasi data.
3. *Transformation* merupakan penyesuaian data manakala terjadi integrasi data dari beberapa sumber agar sesuai dengan target *data warehouse*.
4. *Loading* merupakan pemuatan data ke dalam target *data warehouse*.

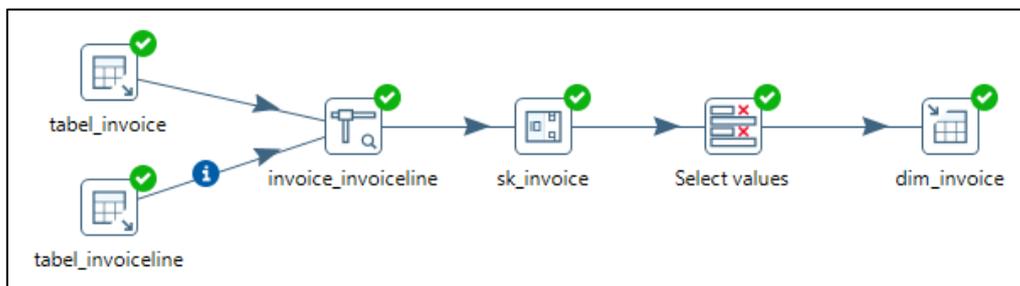
Proses integrasi data pada penelitian ini menggunakan aplikasi *Pentaho Data Integration (PDI)*. Komponen utama dari PDI ini adalah mesin integrasi data berupa perangkat lunak yang mampu menginterpretasi dan mengeksekusi suatu tugas. Tipe objek yang dipergunakan yaitu berupa *transformation* dimana transformasi ini bersifat *data-oriented* dan digunakan untuk mengekstraksi, mentransformasikan, dan memuat data. Transformasi berisi sekumpulan langkah (*steps*), dimana setiap *step* merupakan suatu operasi pada satu atau beberapa *record streams*. Dari satu *step* ke *step* lainnya dihubungkan oleh penghubung yang disebut sebagai *hop*. Suatu *hop* diilustrasikan sebagai sebuah pipa penghubung yang akan mengalirkan record dari satu *step* ke *step* lainnya.

Proses ETL pada penelitian ini arsitektur *data warehouse* menggunakan *Single DDS (DDS Tunggal)* karena aliran data sederhana yang hanya terdiri dari dua data store yaitu *stage* dan *DDS*.



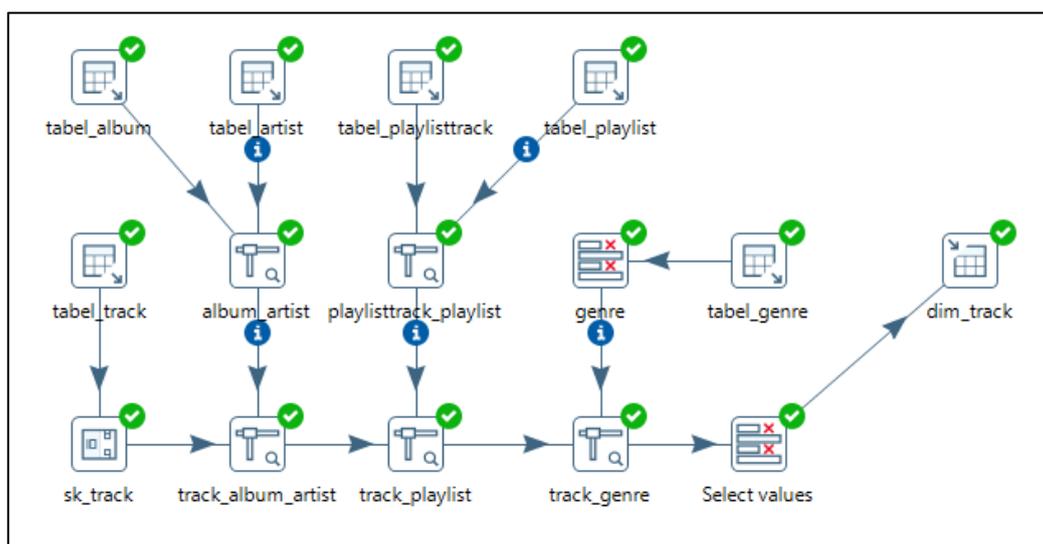
Gambar 15. Proses ETL Tabel Dimensi Customer

Skema proses ETL pada tabel *dim_customer* digambarkan pada gambar 15. *Input* transformasi ini mendapatkan data dari tabel sumber dari *database* yaitu tabel *customer*. Lalu dilakukan *step concat* untuk menggabungkan *field* *FirstName* dan *LastName*. Kemudian ditambahkan *step add sequence* untuk membuat *field* *sk_customer*. Setelah itu dilanjutkan dengan *step select values* untuk menata *field* hasil proses sebelumnya, pemilihan *field* yang akan disimpan serta pemberian nama *field* sesuai dengan tabel *dim_customer*. Terakhir, melakukan *step tabel output* untuk proses *load* dengan nama tabel *dim_customer* pada *database* baru yaitu *dw_chinook*.



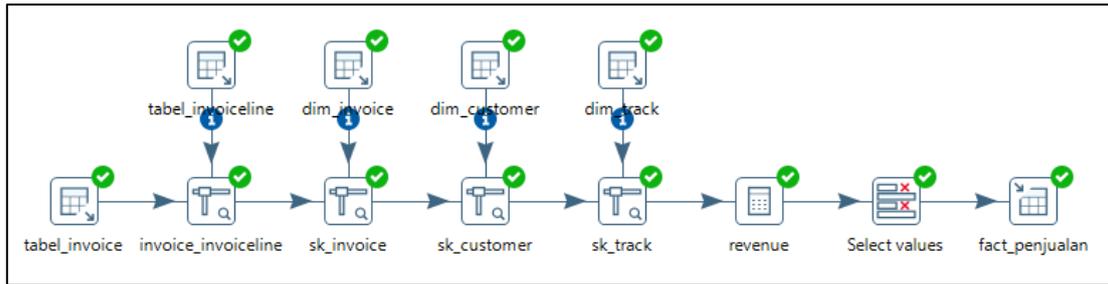
Gambar 16. Proses ETL Tabel Dimensi Invoice

Skema proses ETL pada tabel *dim_invoice* digambarkan pada gambar 16. *Input* transformasi ini mendapatkan data dari dua tabel sumber dari *database* yaitu tabel *invoice* dan tabel *invoiceline*. Lalu dilakukan *step stream lookup* untuk menggabungkan proses *join* pada kedua tabel. Kemudian ditambahkan *step add sequence* untuk membuat *field* *sk_invoice*. Setelah itu dilanjutkan dengan *step select values* untuk menata *field* hasil proses sebelumnya, pemilihan *field* yang akan disimpan serta pemberian nama *field* sesuai dengan tabel *dim_invoice*. Terakhir, melakukan *step tabel output* untuk proses *load* dengan nama tabel *dim_invoice* pada *database* baru yaitu *dw_chinook*.



Gambar 17. Proses ETL Tabel Dimensi Track

Skema proses ETL pada tabel *dim_track* digambarkan pada gambar 17. *Input* transformasi ini mendapatkan data dari beberapa tabel sumber dari *database* yaitu tabel *album*, tabel *artist*, tabel *track*, tabel *playlisttrack*, tabel *playlist* dan tabel *genre*. Lalu dilakukan beberapa *step stream lookup* untuk menggabungkan proses *join* tabel *album* dan *artist* dan juga tabel *playlisttrack* dan tabel *playlist*. Selain itu tabel *genre* dilakukan *step select values* untuk mendapatkan *field* *genre*. Kemudian ditambahkan *step add sequence* untuk membuat *field* *sk_track*. Ketika semua sudah disiapkan maka bisa dilakukan beberapa kali *step stream lookup* untuk menggabungkan semua *field*. Setelah itu dilanjutkan dengan *step select values* untuk menata *field* hasil proses sebelumnya, pemilihan *field* yang akan disimpan serta pemberian nama *field* sesuai dengan tabel *dim_track*. Terakhir, melakukan *step tabel output* untuk proses *load* dengan nama tabel *dim_track* pada *database* baru yaitu *dw_chinook*.



Gambar 18. Proses ETL Tabel Fakta Penjualan

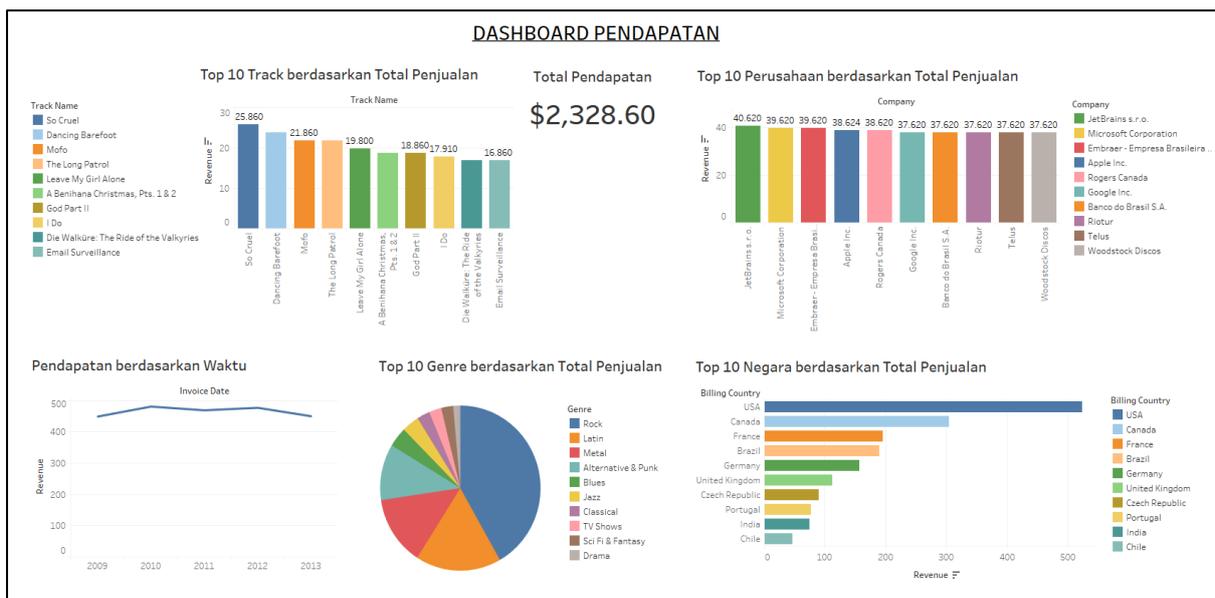
Setelah beberapa tabel dimensi sudah selesai, sekarang untuk tabel fact_penjualan dilakukan *stream lookup* dari beberapa tabel dimensi tersebut. Dengan bantuan hasil *join* dari tabel invoice dan invoiceline maka dapat dilakukan proses *stream lookup* karena *field* pada tabel tersebut bisa digunakan sebagai *lookup field*. Lalu dilanjutkan dengan ditambahkan *step calculator* untuk mendapatkan *field* revenue yang merupakan kalkulasi hasil perkalian dari *field* unit price dan quantity dari penjualan. Setelah itu dilanjutkan dengan *step select values* untuk menata *field* hasil proses sebelumnya, pemilihan *field* yang akan disimpan serta pemberian nama *field* sesuai dengan tabel fact_penjualan. Terakhir, melakukan *step tabel output* untuk proses *load* dengan nama tabel fact_penjualan pada database baru yaitu dw_chinook.

3.4. Penerapan Aplikasi Business Intelligence

Hasil dari *data warehouse* yang dibangun akan divisualisasikan menggunakan aplikasi *business intelligence*. Pada penelitian ini akan digunakan aplikasi Tableau yang merupakan aplikasi *analytic* untuk menganalisa data. Tableau dapat digunakan sebagai tool visualisasi data yang terdapat pada database ataupun *data warehouse*. Untuk melakukan proses visualisasi, akan dilakukan proses pengambilan data dari sumber data. Pada penelitian ini digunakan MySQL database sebagai sumber data. Setelah data berhasil dimuat, maka proses visualisasi data dapat dilakukan.

Visualisasi data yang digunakan pada penelitian ini ialah berupa teks dan grafik, yang akan ditampilkan dalam bentuk *dashboard*. Pada penelitian ini dibuat satu buah *dashboard*, yaitu *dashboard* pendapatan yang dapat dilihat pada gambar 19.

Gambar 19 memuat *dashboard* pendapatan yang memiliki informasi mengenai jumlah pendapatan yang diperoleh oleh perusahaan. Informasi yang disajikan pada dashboard pendapatan ini ialah informasi tentang 10 track dengan penjualan teratas, 10 negara sebagai tujuan pengiriman teratas berdasarkan total pendapatan, 10 perusahaan dengan total pembelian teratas, total penjualan yang didapatkan pada setiap genre, dan total pendapatan berdasarkan waktu.



Gambar 19. Dashboard Pendapatan

4. Kesimpulan

Berdasarkan hasil penelitian mengenai analisis desain dan implementasi *data warehouse* penjualan pada *Chinook Sample Database*, maka penulis dapat mengambil kesimpulan sebagai berikut:

- 1) Pengumpulan data yang digunakan pada penelitian ini didapatkan dengan cara sekunder dengan tahapan studi literatur untuk mendapatkan teori sebagai acuan pembuatan penelitian dan untuk data yang diujikan menggunakan *Chinook Sample Database* yang didapatkan dari situs *yugabyteDB*.
- 2) Proses integrasi *data warehouse* menggunakan konsep ETL (*Extract, Transformation, Load*) dengan bantuan Pentaho Data Integration (PDI).
- 3) Setelah melakukan integrasi data, analisis dilakukan dengan menggunakan query untuk mendapatkan informasi dari database *data warehouse* yang telah diimplementasikan.
- 4) Desain dan implementasi *data warehouse* yang telah dilakukan dapat membantu perusahaan dalam proses pencarian data dan pengambilan keputusan yang terkait dengan penjualan.
- 5) *Data warehouse* tidak hanya menjadi gudang data namun dapat digunakan sebagai pembuatan suatu laporan analisis data dengan efisien.

Referensi

- [1] I. P. A. E. Pratama, *Handbook Data Warehouse Teori Dan Praktik Berbasis Open Source*. Bandung: Informatika Bandung, 2018.
- [2] M. F. Zulkarnain, N. P. N. Ardiyanti, I. W. W. K. Sandi, I. D. N. T. Hendrawan, I. B. M. Mahendra, "Perancangan dan Implementasi Data Warehouse Penjualan (Studi Kasus: Northwind Sample Database)" *Jurnal Elektronik Ilmu Komputer Udayana (JELIKU)*, vol. 10, no. 1, p. 175, 2021.
- [3] Pramatha, C. and N.P.S.H. Mimba, *Udayana University International Student Management: A Business Process Reengineering Approach*. ComTech: Computer, Mathematics and Engineering Applications, 2020. 11(2): p. 57-64.
- [4] D. Subuh and W. Yasman, "Implementasi Data warehouse Dan Penerapannya Pada Toko Magnifique Clothes Dengan Menggunakan Tools Pentaho," *Pros. SENIATI*, p. 29–36, 2019.
- [5] I. P. I. S. Jayadi, M. R. Hartani, W. Y. Astira, P. R. Andrean, I. B. M. Mahendra, "Implementasi Data Warehouse Menggunakan Pentaho BI di Hartaning House Homestay" *Jurnal Elektronik Ilmu Komputer Udayana (JELIKU)*, vol. 10, no. 1, p. 153-161, 2021.
- [6] A. Syam and A. R. Manga, "Sistem Tracer Study Alumni Fakultas Ilmu Komputer Universitas Muslim Indonesia Menggunakan Metode On-Line Analytical Processing (OLAP)," *Ilk. J. Ilm.*, vol. 9, no. 1, p. 86–90, 2018.

This page is intentionally left blank