

Dibimbing Business Intelligence Bootcamp
Assignment Day 19
BI Case Study 01 – Data Warehouse

Name : Dartwin

Email : dartwinshu123@gmail.com

1. Menggunakan data berikut pilih bisnis use case yang akan teman-teman kerjakan untuk membangun data warehouse. Apakah teman-teman akan fokus terhadap transaksi? Ataukah customer?

Jawab:

Dengan menjalankan SQL script restore dan meng-copy data yang sudah ada ke dalam PostgreSQL untuk mengetahui data apa saja yang disediakan, bisnis use case yang dipilih adalah **membuat data warehouse untuk menampung/menyimpan data transaksi rental DVD yang dilakukan oleh customer**, dimana data warehouse yang dirancang akan berfokus kepada customer, karena industri rental DVD adalah industri yang sangat bergantung pada customer-nya. Sebab industri rental DVD memenuhi kebutuhan sekunder, bahkan tersier dari customer, berbeda jauh dengan industri jasa umum seperti ride-hailing, waste-management, dsb..

Oleh karena itu, pelaku bisnis dalam industri ini lebih jeli dan peka terhadap behavior transaksi customer sehingga dapat menentukan strategi marketing untuk mendapatkan customer baru dan customer rewarding program untuk mempertahankan pelanggan, bahkan menentukan business model baru yang sustainable.

Bisnis case yang ditentukan akan membangun data warehouse yang dapat support dalam pembuatan/proses:

- Data analytics
- Machine learning (baik untuk classification, maupun clustering)
- Dashboard

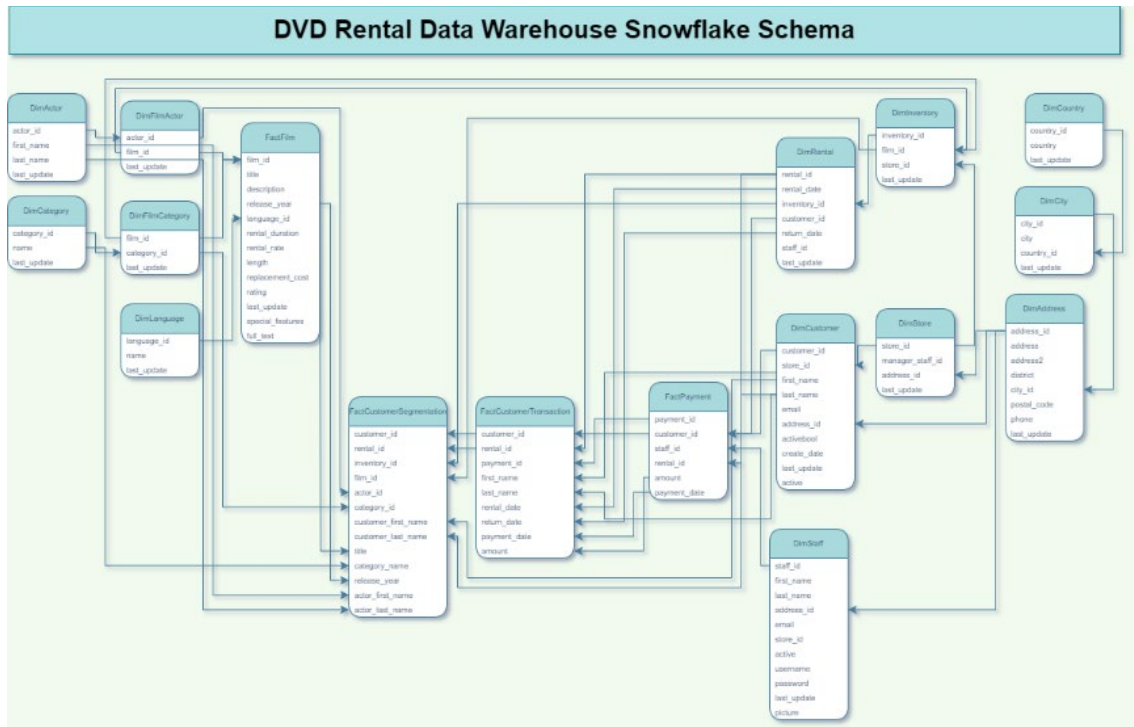
sehingga dapat membantu BoD atau managerial level perusahaan melakukan:

- Strategi marketing
- Customer rewarding
- Customer segmentation
- Demand forecasting (Trend analysis)
- Etc.

- Buatlah schema data warehouse dengan tepat. Tidak lupa juga susun fact dan dimension table yang dibutuhkan menggunakan star schema atau snowflake schema!

Jawab:

Terlampir gambar DVD rental data warehouse snowflake schema yang dibuat. Akan dilampiri juga gambar DVD rental data warehouse snowflake melalui PDF sehingga lebih mudah untuk dipahami.



- Jumlah table yang dibuat juga menjadi penilaian buatlah sekitar 2 table untuk data warehouse dan 4 table untuk data mart.

Jawab:

1. Tabel Data Warehouse: Customer Segmentation

	customer_id	rental_id	inventory_id	film_id	actor_id	category_id	customer_name	title	category_name	release_year	actor_name
1	375	1,499	165	35	7	11	Aaron Selby	Arachnophobia Rollercoaster	Horror	2,006	Grace Mosele
2	375	1,499	165	35	184	11	Aaron Selby	Arachnophobia Rollercoaster	Horror	2,006	Humphrey Garland
3	375	1,499	165	35	139	11	Aaron Selby	Arachnophobia Rollercoaster	Horror	2,006	Ewan Gooding
4	375	1,499	165	35	135	11	Aaron Selby	Arachnophobia Rollercoaster	Horror	2,006	Rita Reynolds
5	375	1,499	165	35	118	11	Aaron Selby	Arachnophobia Rollercoaster	Horror	2,006	Cuba Allen
6	375	1,499	165	35	113	11	Aaron Selby	Arachnophobia Rollercoaster	Horror	2,006	Morgan Hopkins
7	375	1,499	165	35	95	11	Aaron Selby	Arachnophobia Rollercoaster	Horror	2,006	Daryl Wahlberg
8	375	1,499	165	35	35	11	Aaron Selby	Arachnophobia Rollercoaster	Horror	2,006	Judy Dean
9	375	10,859	277	61	174	7	Aaron Selby	Beauty Grease	Drama	2,006	Michael Bening
10	375	10,859	277	61	111	7	Aaron Selby	Beauty Grease	Drama	2,006	Cameron Zellweger
11	375	10,859	277	61	136	7	Aaron Selby	Beauty Grease	Drama	2,006	Ed Mansfield
12	375	10,859	277	61	138	7	Aaron Selby	Beauty Grease	Drama	2,006	Lucille Dee
13	375	10,859	277	61	166	7	Aaron Selby	Beauty Grease	Drama	2,006	Nick Degeneres
14	375	3,981	842	184	170	4	Aaron Selby	Core Suit	Classics	2,006	Mena Hopper
15	375	3,981	842	184	54	4	Aaron Selby	Core Suit	Classics	2,006	Penelope Pinkett
16	375	3,981	842	184	104	4	Aaron Selby	Core Suit	Classics	2,006	Penelope Cromyn
17	375	3,981	842	184	144	4	Aaron Selby	Core Suit	Classics	2,006	Angela Witherspoon
18	375	15,505	1,144	253	14	1	Aaron Selby	Drifter Commandments	Action	2,006	Vivien Bergen
19	375	15,505	1,144	253	37	1	Aaron Selby	Drifter Commandments	Action	2,006	Val Bolger
20	375	15,505	1,144	253	60	1	Aaron Selby	Drifter Commandments	Action	2,006	Henry Berry

2. Tabel Data Warehouse: Customer Transaction

	customer_id	rental_id	payment_id	customer_name	rental_date	return_date	payment_date	amount
1	13	8,831	29,136	Karen Jackson	2005-07-29 22:37:41.000	2005-08-08 04:28:41.000	2007-04-29 21:06:07.996	11.99
2	591	4,383	28,799	Kent Arsenault	2005-07-07 20:45:51.000	2005-07-17 00:52:51.000	2007-04-07 19:14:17.996	11.99
3	362	14,759	20,403	Nicholas Barfield	2005-08-21 23:28:58.000	2005-08-31 00:35:58.000	2007-03-21 21:57:24.996	11.99
4	204	15,415	22,650	Rosemary Schmidt	2005-08-22 23:48:56.000	2005-09-01 02:05:56.000	2007-03-22 22:17:22.996	11.99
5	237	11,479	24,866	Tanya Gilbert	2005-08-02 22:18:13.000	2005-08-12 00:43:13.000	2007-03-02 20:46:39.996	11.99
6	592	3,973	28,814	Terrance Roush	2005-07-06 22:58:31.000	2005-07-16 03:27:31.000	2007-04-06 21:26:57.996	11.99
7	195	16,040	24,553	Vanessa Sims	2005-08-23 22:19:33.000	2005-09-02 02:19:33.000	2007-03-23 20:47:59.996	11.99
8	116	14,763	23,757	Victoria Gibson	2005-08-21 23:34:00.000	2005-08-31 03:58:00.000	2007-03-21 22:02:26.996	11.99
9	367	14,481	20,449	Adam Gooch	2005-08-21 13:41:14.000	2005-08-30 07:53:14.000	2007-03-21 12:09:40.996	10.99
10	567	15,594	22,352	Alfredo Mcadams	2005-08-23 06:18:43.000	2005-09-01 09:09:43.000	2007-03-23 04:47:09.996	10.99
11	196	10,122	31,206	Alma Austin	2005-07-31 21:29:28.000	2005-08-09 00:04:28.000	2007-04-30 19:57:54.996	10.99
12	333	9,338	25,921	Andrew Purdy	2005-07-30 18:03:13.000	2005-08-08 23:57:13.000	2007-04-30 16:31:39.996	10.99
13	29	5,442	29,300	Angela Hernandez	2005-07-09 21:55:19.000	2005-07-19 01:18:19.000	2007-04-09 20:23:45.996	10.99
14	136	11,992	23,952	Anita Morales	2005-08-17 18:27:22.000	2005-08-26 22:38:22.000	2007-03-17 16:55:48.996	10.99
15	33	1,301	18,640	Anna Hill	2005-06-15 09:46:33.000	2005-06-24 08:56:33.000	2007-02-15 08:14:59.996	10.99
16	364	5,872	26,280	Benjamin Varney	2005-07-10 18:54:05.000	2005-07-19 15:41:05.000	2007-04-10 17:22:31.996	10.99
17	221	2,660	19,336	Bessie Morrison	2005-06-19 10:50:02.000	2005-06-28 12:49:02.000	2007-02-19 09:18:28.996	10.99
18	457	14,327	21,327	Bill Gavin	2005-08-21 08:18:18.000	2005-08-30 02:21:18.000	2007-03-21 06:46:44.996	10.99
19	371	11,086	20,480	Billy Poulin	2005-08-02 07:38:44.000	2005-08-11 06:30:44.000	2007-03-02 06:07:10.996	10.99
20	564	7,470	28,479	Bob Pfeiffer	2005-07-27 19:01:03.000	2005-08-05 19:16:03.000	2007-04-27 17:29:29.996	10.99

3. Tabel Data Mart: Jumlah Customer untuk Setiap Kategori Film

	category_name	jumlah_customer
1	Sports	6,398
2	Animation	6,052
3	Action	5,736
4	Documentary	5,521
5	Drama	5,347
6	Sci-Fi	5,233
7	Foreign	5,135
8	Family	5,050
9	Children	5,019
10	New	4,757
11	Classics	4,663
12	Horror	4,520
13	Travel	4,300
14	Comedy	4,185
15	Games	4,161
16	Music	4,076

4. Tabel Data Mart: Jumlah Customer untuk Setiap Actor

	asc actor_name	123 jumlah_customer
1	Susan Davis	750
2	Gina Degeneres	685
3	Matthew Carrey	619
4	Mary Keitel	607
5	Angela Witherspoon	589
6	Walter Torn	584
7	Henry Berry	564
8	Jayne Nolte	564
9	Val Bolger	562
10	Sandra Kilmer	556
11	Sean Guinness	538
12	Ewan Gooding	523
13	Scarlett Damon	519
14	Angela Hudson	516
15	Kevin Garland	511
16	Warren Nolte	511
17	Helen Voight	510
18	Woody Hoffman	508
19	Cameron Zellweger	507
20	Michael Bolger	504

5. Tabel Data Mart: Top 10 Film dengan Jumlah Customer Terbanyak

	asc title	125 jumlah_customer
1	Telemark Heartbreakers	264
2	Titanic Boondock	264
3	Juggler Hardly	261
4	Dracula Crystal	247
5	Oz Liaisons	242
6	Family Sweet	234
7	Saturday Lambs	234
8	Hellfighters Sierra	228
9	Suspects Quills	216
10	Coneheads Smoochy	216

6. Tabel Data Mart: Top 100 Customer yang Melakukan Transaksi dengan Nilai Terbesar

	max customer_name	123 jumlah_nilai_transaksi
1	Eleanor Hunt	211.55
2	Karl Seal	208.58
3	Marion Snyder	194.61
4	Rhonda Kennedy	191.62
5	Clara Shaw	189.6
6	Tommy Collazo	183.63
7	Ana Bradley	167.67
8	Curtis Irby	167.62
9	Marcia Dean	166.61
10	Mike Way	162.67
11	Arnold Havens	161.68
12	Wesley Bull	158.65
13	Gordon Allard	157.69
14	Louis Leone	156.66
15	Lena Jensen	154.7
16	Tim Cary	154.66
17	Warren Sherrod	152.69
18	Steve Mackenzie	152.68
19	Brittany Riley	151.73
20	Guy Brownlee	151.69

4. Implementasi dalam bentuk SQL untuk data warehouse dan data mart yang telah dirancang.

Jawab:

SQL Code

```
-- Pembuatan Tabel Data Warehouse Customer Segmentation
```

```
SELECT
```

```
    c.customer_id,
```

```
    r.rental_id,
```

```
    r.inventory_id,
```

```
    f.film_id,
```

```
    a.actor_id,
```

```
    ctg.category_id,
```

```
    c.first_name || ' ' || c.last_name AS customer_name,
```

```
    f.title,
```

```
    ctg.name AS category_name,
```

```
    f.release_year,
```

```
    a.first_name || ' ' || a.last_name AS actor_name
```

```
FROM customer c
```

```
    LEFT JOIN
```

```
    payment p
```

```
ON c.customer_id = p.customer_id
```

```
    LEFT JOIN
```

```

        rental r
    ON p.rental_id = r.rental_id
        LEFT JOIN
        inventory i
    ON r.inventory_id = i.inventory_id
        LEFT JOIN
        film f
    ON i.film_id = f.film_id
        LEFT JOIN
        film_actor fa
    ON f.film_id = fa.film_id
        LEFT JOIN
        actor a
    ON fa.actor_id = a.actor_id
        LEFT JOIN
        film_category fctg
    ON f.film_id = fctg.film_id
        LEFT JOIN
        category ctg
    ON fctg.category_id = ctg.category_id
    ORDER BY customer_name ASC, f.title ASC, f.release_year DESC;

```

-- Pembuatan Tabel Data Warehouse Customer Transaction

```

SELECT
    c.customer_id,
    r.rental_id,
    p.payment_id,
    c.first_name || ' ' || c.last_name AS customer_name,
    r.rental_date,
    r.return_date,
    p.payment_date,
    p.amount
FROM customer c
    LEFT JOIN
    payment p
    ON c.customer_id = p.customer_id
        LEFT JOIN
        rental r
    ON p.rental_id = r.rental_id
    ORDER BY p.amount DESC, customer_name ASC;

```

-- Pembuatan Tabel Data Mart Jumlah Customer untuk Setiap Kategori Film

```

SELECT
    ctg.name AS category_name,
    count(c.customer_id) AS jumlah_customer

```

```

FROM customer c
    LEFT JOIN
    payment p
ON c.customer_id = p.customer_id
    LEFT JOIN
    rental r
ON p.rental_id = r.rental_id
    LEFT JOIN
    inventory i
ON r.inventory_id = i.inventory_id
    LEFT JOIN
    film f
ON i.film_id = f.film_id
    LEFT JOIN
    film_actor fa
ON f.film_id = fa.film_id
    LEFT JOIN
    actor a
ON fa.actor_id = a.actor_id
    LEFT JOIN
    film_category fctg
ON f.film_id = fctg.film_id
    LEFT JOIN
    category ctg
ON fctg.category_id = ctg.category_id
GROUP BY ctg.name
ORDER BY jumlah_customer DESC;

```

-- Pembuatan Tabel Data Mart Jumlah Customer untuk Setiap Actor

```

SELECT
    a.first_name || ' ' || a.last_name AS actor_name,
    count(c.customer_id) AS jumlah_customer
FROM customer c
    LEFT JOIN
    payment p
ON c.customer_id = p.customer_id
    LEFT JOIN
    rental r
ON p.rental_id = r.rental_id
    LEFT JOIN
    inventory i
ON r.inventory_id = i.inventory_id
    LEFT JOIN
    film f
ON i.film_id = f.film_id
    LEFT JOIN

```

```

        film_actor fa
    ON f.film_id = fa.film_id
    LEFT JOIN
        actor a
    ON fa.actor_id = a.actor_id
    LEFT JOIN
        film_category fctg
    ON f.film_id = fctg.film_id
    LEFT JOIN
        category ctg
    ON fctg.category_id = ctg.category_id
GROUP BY actor_name
ORDER BY jumlah_customer DESC;

```

-- Pembuatan Tabel Data Mart Top 10 Film Jumlah Customer Terbanyak

```

SELECT
    f.title,
    count(c.customer_id) AS jumlah_customer
FROM customer c
    LEFT JOIN
        payment p
    ON c.customer_id = p.customer_id
    LEFT JOIN
        rental r
    ON p.rental_id = r.rental_id
    LEFT JOIN
        inventory i
    ON r.inventory_id = i.inventory_id
    LEFT JOIN
        film f
    ON i.film_id = f.film_id
    LEFT JOIN
        film_actor fa
    ON f.film_id = fa.film_id
    LEFT JOIN
        actor a
    ON fa.actor_id = a.actor_id
    LEFT JOIN
        film_category fctg
    ON f.film_id = fctg.film_id
    LEFT JOIN
        category ctg
    ON fctg.category_id = ctg.category_id
GROUP BY f.title
ORDER BY jumlah_customer DESC
LIMIT 10;

```


-- Pembuatan Tabel Data Mart top 100 Customer yang Melakukan Transaksi dengan Nilai Terbesar

```
SELECT DISTINCT
    c.first_name || ' ' || c.last_name AS customer_name,
    sum(p.amount) OVER (PARTITION BY c.first_name || ' ' || c.last_name ORDER BY
        c.first_name || ' ' || c.last_name) AS "jumlah_nilai_transaksi"
FROM customer c
    LEFT JOIN
        payment p
ON c.customer_id = p.customer_id
    LEFT JOIN
        rental r
ON p.rental_id = r.rental_id
ORDER BY jumlah_nilai_transaksi DESC, customer_name ASC
LIMIT 100;
```