Data Warehouse (Reporting Layer)

17-12-2024 by DbSchema.com - Wise Coders

Layouts

Tables

Views

Since all the definintion are given in the curated layer,
here we have definitions for the tables and columns which are new or in a derived form from the curated layer

disease_type_dimension
- disease_type_cd
- disease_type_description t

disease_type_cd

disease_dimension
- disease_id
- disease_name t
- disease_type_cd
- intensity_level #

manufacturer_dimension
- manufacturer_surrogate_key
- manufacturer_id #
- manufacturer_name t
- effective_start_date d
- effective_end_date d
- is_current_flag t

race_gender_dimension
- combination_id
- race_code t
- race_description t
- gender t

combination_id

disease_id

disease_id

manufacturer_id

diseased_patient_fact
- disease_id
- location_id
- combination_id
- num_of_patients #
- race_propensity_value #
- avg_severity_value #
- start_of_month_date

medicine_fact
- fact_id #
- disease_id
- first_available_date
- manufacturer_id
- num_of_medicines #
- avg_effectiveness_of_medicines #

location_dimension
- location_surrogate_key
- location_id #
- city_name t
- state_province_name t
- country_name t
- developing_flag t
- wealth_rank_number #
- effective_start_date d
- effective_end_date d
- is_current_flag t

location_id

month_date

vailable_date

date_dimension
- date_key #
- full_date
- day #
- month #
- year #
- quarter #
- week #
- day_name t
- month_name t

Views

manufacturer_medicine_effectiveness
- manufacturer_name t
- total_medicines_produced #
- avg_medicine_effectiveness #

medicine_effectiveness_by_disease
- disease_name t
- manufacturer_name t
- total_medicines #
- avg_effectiveness #

disease_distribution_by_gender_race
- disease_name t
- race_description t
- gender t
- num_of_reports #

monthly_disease_severity_trends
- disease_name t
- intensity_level #
- disease_type_cd t
- month d
- avg_severity #

disease_statistics_by_location
- disease_name t
- city_name t
- state_province_name t
- country_name t
- total_patients #
- avg_severity #
- avg_propagation #

active_locations_by_disease
- disease_name t
- city_name t
- state_province_name t
- country_name t
- num_of_reports #

Main Layout

Table date_dimension

Idx	Name	Data Type	Description
The table will enable us to slice and dice the data with various date attributes like quarter, week etc. to aid the analysis.			
* Pk	date_key	integer	Unique key for the Date
* Unq	full_date	date	The full date in yyyy-mm-dd format
	day	integer	Day of the date
	month	integer	Month of the date
	year	integer	Year of the date
	quarter	integer	Quarter of the date
	week	integer	Week of the date
	day_name	varchar(20)	Day name of the date
	month_name	varchar(20)	Month name of the date

Indexes

Type	Name	On	Description
Pk	date_dimension_pkey	date_key	
Unq	date_dimension_full_date_key	full_date	

Table disease_dimension

Idx	Name	Data Type
* Pk	disease_id	integer
*	disease_name	varchar(100)
*	disease_type_cd	varchar(5)
	intensity_level	integer

Indexes

Type	Name	On
Pk	disease_dimension_pkey	disease_id

Foreign Keys

Type	Name	On
	fk_disease_type (disease_type_cd) ref disease_type_dimension (disease_type_cd)	

Constraints

	Name	Definition
	disease_dimension_intensity_level_check	((intensity_level >= 1) AND (intensity_level <= 10))

Table disease_type_dimension

Idx	Name	Data Type

Table disease_type_dimension

* Pk	disease_type_cd	varchar(5)
	disease_type_description	varchar(100)

Indexes

Type	Name	On
Pk	disease_type_dimension_pkey	disease_type_cd

Table diseased_patient_fact

Idx	Name	Data Type	Description
* Pk	disease_id	integer	
* Pk	location_id	integer	
* Pk	combination_id	integer	
*	num_of_patients	integer	Count of number of patients
	race_propensity_value	numeric(5,2)	
	avg_severity_value	numeric(5,2)	Avg Severity based based on all the patients under this particular bucket. (value will be between 1 an 10)
* Pk	start_of_month_date	date	The date of the start of the month when the patient got the disease. (Aggregation)

Indexes

Type	Name	On	Description
Pk	diseased_patient_fact_pkey	disease_id, location_id, combination_id, start_of_month_date	

Foreign Keys

Type	Name	On	Description
	diseased_patient_fact_start_of_month_date_fkey (start_of_month_date) ref date_dimension (full_date)		
	diseased_patient_fact_combination_id_fkey (combination_id) ref race_gender_dimension (combination_id)		
	diseased_patient_fact_disease_id_fkey (disease_id) ref disease_dimension (disease_id)		
	diseased_patient_fact_location_id_fkey (location_id) ref location_dimension (location_surrogate_key)		

Table location_dimension

Idx	Name	Data Type	Description
The table has the SDC type 2 for the location IDs as the names can change over time.			
* Pk	location_surrogate_key	integer DEFAULT nextval('reporting_layer_dw.location_dimension_location_surrogate_key_seq'::regclass)	
* Unq	location_id	integer	

Table location_dimension

	city_name	varchar(100)	
	state_province_name	varchar(100)	
	country_name	varchar(100)	
	developing_flag	varchar(1)	
	wealth_rank_number	integer	
Unq	effective_start_date	date DEFAULT '1899-01-01'::date	
Unq	effective_end_date	date	
	is_current_flag	varchar(1)	

Indexes

Type	Name	On	Description
Pk	location_dimension_pkey	location_surrogate_key	
Unq	unique_location_id_effective_start_end_date	location_id, effective_start_date, effective_end_date	

Constraints

	Name	Definition	Description
	location_dimension_developing_flag_check	((developing_flag)::text = ANY ((ARRAY['Y'::character varying, 'N'::character varying])::text[]))	
	location_dimension_wealth_rank_number_check	((wealth_rank_number >= 1) AND (wealth_rank_number <= 10))	
	location_dimension_is_current_flag_check	((is_current_flag)::text = ANY ((ARRAY['Y'::character varying, 'N'::character varying])::text[]))	

Table manufacturer_dimension

Idx	Name	Data Type	Description
This dimension will capture the Pharmaceutical company information. Follows SDC type 2 to handle mergers and bankruptcy			
* Pk	manufacturer_surrogate_key	integer DEFAULT nextval('reporting_layer_dw.manufacturer_dimension_manufacturer_surrogate_key_seq'::regclass)	
* Unq	manufacturer_id	integer	
*	manufacturer_name	varchar(150)	
Unq	effective_start_date	date DEFAULT '1899-01-01'::date	
Unq	effective_end_date	date	
	is_current_flag	varchar(1)	

Indexes

Type	Name	On	Description
Pk	manufacturer_dimension_pkey	manufacturer_surrogate_key	

Table manufacturer_dimension

Unq	unique_manufacturer_id_effective_start_end_date	manufacturer_id, effective_start_date, effective_end_date

Constraints

Name	Definition	Description
manufacturer_dimension_is_current_flag_check	((is_current_flag)::text = ANY ((ARRAY['Y'::character varying, 'N'::character varying])::text[]))	

Table medicine_fact

Idx	Name	Data Type	Description
* Pk	fact_id	integer DEFAULT nextval('reporting_layer_dw.medicine_fact_fact_id_seq'::regclass)	
*	disease_id	integer	
*	first_available_date	date	
	manufacturer_id	integer	
	num_of_medicines	integer	Number of Medicines manufactured by the Manufacturer for a given disease
	avg_effectiveness_of_medicines	numeric(5,2)	Average effectiveness of all the medicines of a particular manufacturer for a disease

Indexes

Type	Name	On	Description
Pk	medicine_fact_pkey	fact_id	

Foreign Keys

Type	Name	On	Description
	medicine_fact_first_available_date_fkey (first_available_date) ref date_dimension (full_date)		
	medicine_fact_manufacturer_id_fkey (manufacturer_id) ref manufacturer_dimension (manufacturer_surrogate_key)		
	medicine_fact_disease_id_fkey (disease_id) ref disease_dimension (disease_id)		

Table race_gender_dimension

Idx	Name	Data Type
* Pk	combination_id	integer
*	race_code	varchar(5)
	race_description	varchar(100)
	gender	varchar(1)

Indexes

Type	Name	On
Pk	race_gender_dimension_pkey	combination_id

Table race_gender_dimension

Constraints

Name	Definition
race_gender_dimension_gender_check	((gender)::text = ANY ((ARRAY['M'::character varying, 'F'::character varying, 'O'::character varying, 'U'::character varying])::text[]))

View active_locations_by_disease

```
CREATE OR REPLACE VIEW ${view} AS SELECT dd.disease_name,
    ld.city_name,
    ld.state_province_name,
    ld.country_name,
    count(dpf.location_id) AS num_of_reports
   FROM ((reporting_layer_dw.diseased_patient_fact dpf
     JOIN reporting_layer_dw.disease_dimension dd ON ((dpf.disease_id = dd.disease_id)))
     JOIN reporting_layer_dw.location_dimension ld ON ((dpf.location_id = ld.location_surrogate_key)))
  GROUP BY dd.disease_name, ld.city_name, ld.state_province_name, ld.country_name
  ORDER BY dd.disease_name, (count(dpf.location_id)) DESC
```

View disease_distribution_by_gender_race

```
CREATE OR REPLACE VIEW ${view} AS SELECT dd.disease_name,
    rgd.race_description,
    rgd.gender,
    count(dpf.combination_id) AS num_of_reports
   FROM ((reporting_layer_dw.diseased_patient_fact dpf
     JOIN reporting_layer_dw.disease_dimension dd ON ((dpf.disease_id = dd.disease_id)))
     JOIN reporting_layer_dw.race_gender_dimension rgd ON ((dpf.combination_id = rgd.combination_id)))
  GROUP BY dd.disease_name, rgd.race_description, rgd.gender
  ORDER BY dd.disease_name, (count(dpf.combination_id)) DESC
```

View disease_statistics_by_location

```
CREATE OR REPLACE VIEW ${view} AS SELECT dd.disease_name,
    ld.city_name,
    ld.state_province_name,
    ld.country_name,
    sum(dpf.num_of_patients) AS total_patients,
    round(avg(dpf.avg_severity_value), 2) AS avg_severity,
    round(avg(dpf.race_propensity_value), 2) AS avg_propagation
   FROM ((reporting_layer_dw.diseased_patient_fact dpf
     JOIN reporting_layer_dw.disease_dimension dd ON ((dpf.disease_id = dd.disease_id)))
     JOIN reporting_layer_dw.location_dimension ld ON ((dpf.location_id = ld.location_surrogate_key)))
  GROUP BY dd.disease_name, ld.city_name, ld.state_province_name, ld.country_name
```

View manufacturer_medicine_effectiveness

```
CREATE OR REPLACE VIEW ${view} AS SELECT md.manufacturer_name,
    sum(mf.num_of_medicines) AS total_medicines_produced,
    round(avg(mf.avg_effectiveness_of_medicines), 2) AS avg_medicine_effectiveness
   FROM (reporting_layer_dw.medicine_fact mf
     JOIN reporting_layer_dw.manufacturer_dimension md ON ((mf.manufacturer_id = md.manufacturer_surrogate_key)))
  GROUP BY md.manufacturer_name
  ORDER BY (round(avg(mf.avg_effectiveness_of_medicines), 2)) DESC
```

View medicine_effectiveness_by_disease

```
CREATE OR REPLACE VIEW ${view} AS SELECT dd.disease_name,
    md.manufacturer_name,
    sum(mf.num_of_medicines) AS total_medicines,
    round(avg(mf.avg_effectiveness_of_medicines), 2) AS avg_effectiveness
   FROM ((reporting_layer_dw.medicine_fact mf
     JOIN reporting_layer_dw.disease_dimension dd ON ((mf.disease_id = dd.disease_id)))
     JOIN reporting_layer_dw.manufacturer_dimension md ON ((mf.manufacturer_id = md.manufacturer_surrogate_key)))
  GROUP BY dd.disease_name, md.manufacturer_name
```

View monthly_disease_severity_trends

```
CREATE OR REPLACE VIEW ${view} AS SELECT dd.disease_name,
    dd.intensity_level,
    dd.disease_type_cd,
    date_trunc('month'::text, (dpf.start_of_month_date)::timestamp with time zone) AS month,
    round(avg(dpf.avg_severity_value), 2) AS avg_severity
   FROM (reporting_layer_dw.diseased_patient_fact dpf
     JOIN reporting_layer_dw.disease_dimension dd ON ((dpf.disease_id = dd.disease_id)))
  GROUP BY dd.disease_name, dd.intensity_level, dd.disease_type_cd, (date_trunc('month'::text, (dpf.start_of_month_date)::timestamp with time zone))
  ORDER BY (date_trunc('month'::text, (dpf.start_of_month_date)::timestamp with time zone)), dd.disease_name
```