

Term Project Final Presentation

Pragya Sen, BUID: U97864862

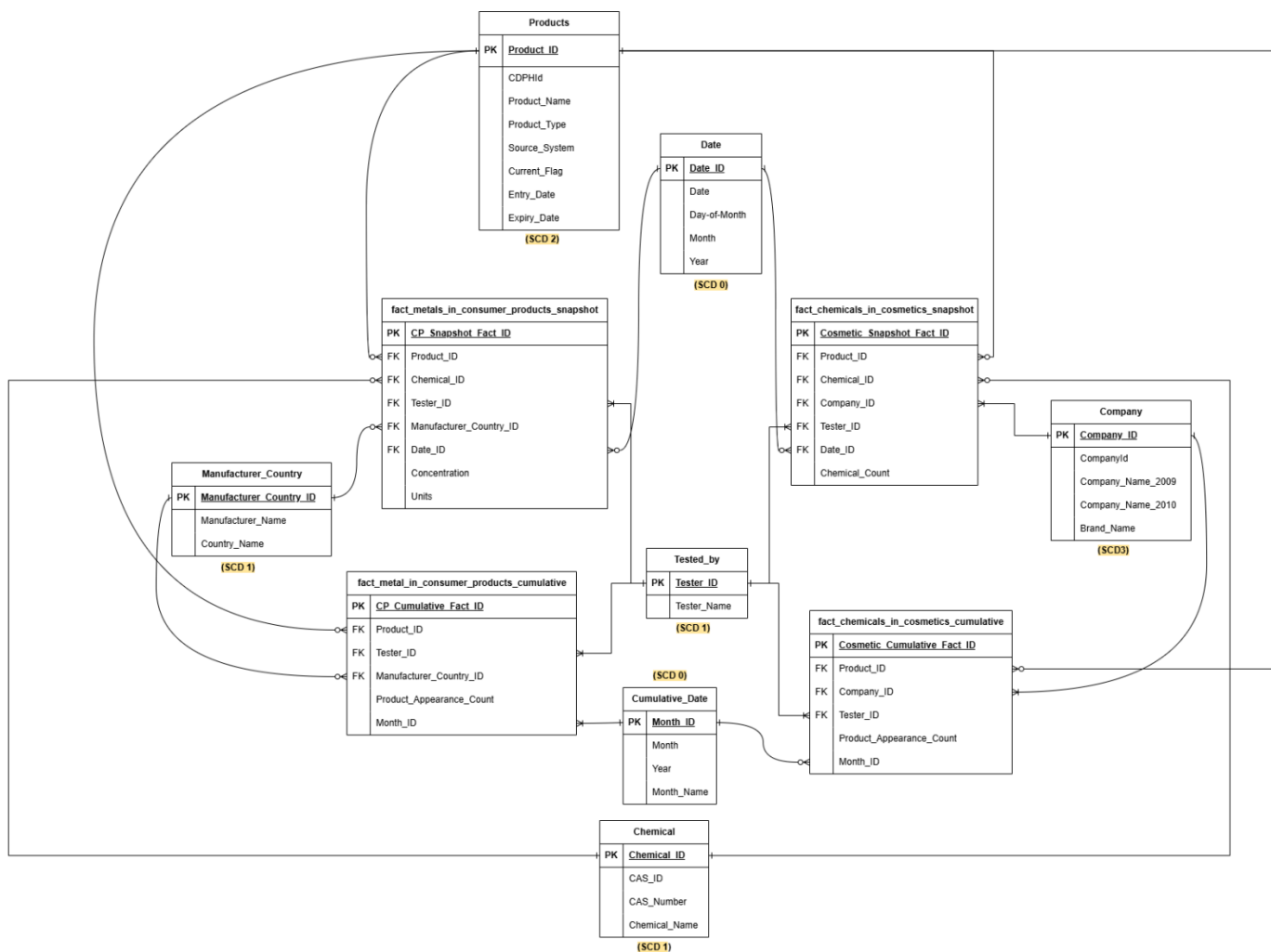
Datasets:

- Chemicals in Cosmetics (from official California state dept.)
- Metals in Consumer products (from official New York state dept.)

For staging Data:

```
\copy staging_consumer(Row_ID, ProductType, ProductName, Metal, Concentration, Units,
Manufacturer, Made_In_Country, Purchase_Country, Collection_Date, Investigation_Type)
FROM 'C:/Users/pragy/OneDrive/Desktop/Spring2025/779_Advanced_DBMS/Term
Project/Metal_Content_of_Consumer_Products_Testing_by_the_NYC_Health_Department.csv'
DELIMITER ';' CSV HEADER ENCODING 'UTF8';
```

ERD Implemented:



Steps:

- Created dimension and fact tables.
- Filled data from datasets into 2 staging tables.
- Moved data from these staging tables to the created dimension and fact tables.
- Checked tables for different SCDs.
- Queries for answering business questions.

Delta Report

1. Handling inserts into SCD2 tables using Stored Procedures.

Using Regular Stored Procedure:

```
Query Query History
220 CREATE OR REPLACE PROCEDURE insert_dim_product_scd2(
221     p_cdph_id INTEGER,
222     p_product_name VARCHAR,
223     p_product_type VARCHAR,
224     p_source_system VARCHAR
225 )
226 LANGUAGE plpgsql
227 AS $$
228 DECLARE
229     existing_record dim_products%ROWTYPE;
230 BEGIN
231     --Checking if current record exists for the same business key (cdph_id,
232     SELECT *
233     INTO existing_record
234     FROM dim_products
235     WHERE cdph_id = p_cdph_id
236           AND source_system = p_source_system
237           AND current_flag = TRUE;
238
239     --If no such record exists, inserting new one
240     IF NOT FOUND THEN
241         INSERT INTO dim_products (cdph_id, product_name, product_type, source_system)
242         VALUES (p_cdph_id, p_product_name, p_product_type, p_source_system);
243     ELSE
244         --Checking if attributes differ
245         IF existing_record.product_name IS DISTINCT FROM p_product_name
```

Query Query History

```

245         IF existing_record.product_name IS DISTINCT FROM p_product_name
246         OR existing_record.product_type IS DISTINCT FROM p_product_type T
247
248         --Expire old record
249         UPDATE dim_products
250         SET current_flag = FALSE,
251           expiry_date = CURRENT_DATE
252         WHERE product_id = existing_record.product_id;
253
254         --Insert new version
255         INSERT INTO dim_products (cdph_id, product_name, product_type, s
256         VALUES (p_cdph_id, p_product_name, p_product_type, p_source_syst
257         END IF;
258         --Else: No change, do nothing
259     END IF;
260 END;
261 $$;
262
263 --testing
264 select * from dim_products limit 1
265 --making changes to existing record
266 CALL insert_dim_product_scd2(
267     7151,
268     'Bluebell Shower Gel',
269     'Bath Products',
270     'COSMETICS'
271 );
272
273 --trying to insert an identical record
274 CALL insert_dim_product_scd2(7151, 'Bluebell Shower Gel', 'Bath Products', 'COSMET

```

Before Insert:

263 --testing
264 select * from dim_products limit 1

product_id [PK] integer	cdph_id integer	product_name character varying (255)	product_type character varying (100)	source_system character varying (20)	current_flag boolean	entry_date date	expiry_date date
1	7151	Bluebell Bath & Shower Gel	Bath Products	COSMETICS	true	2025-04-25	9999-12-31

After Insert:

273 select * from dim_products where cdph_id = 7151

product_id [PK] integer	cdph_id integer	product_name character varying (255)	product_type character varying (100)	source_system character varying (20)	current_flag boolean	entry_date date	expiry_date date
1	7151	Bluebell Bath & Shower Gel	Bath Products	COSMETICS	false	2025-04-25	2025-05-02
2	42198	Bluebell Shower Gel	Bath Products	COSMETICS	true	2025-05-02	9999-12-31

Using MERGE function in Stored Procedure:

```
Query  Query History
278  --using merge function
279  CREATE OR REPLACE PROCEDURE insert_dim_product_scd2_merge(
280      p_cdph_id INTEGER,
281      p_product_name VARCHAR,
282      p_product_type VARCHAR,
283      p_source_system VARCHAR
284  )
285  LANGUAGE plpgsql
286  AS $$
287  DECLARE
288      existing_product_id INTEGER;
289      changes_needed BOOLEAN;
290  BEGIN
291      --Checking if an active version exists
292      SELECT product_id INTO existing_product_id
293      FROM dim_products
294      WHERE cdph_id = p_cdph_id
295            AND source_system = p_source_system
296            AND current_flag = TRUE;
297
298      --If it exists, checking if changes are needed
299      IF existing_product_id IS NOT NULL THEN
300          SELECT (product_name IS DISTINCT FROM p_product_name OR product_type IS DISTINCT FROM p_product_type)
301          INTO changes_needed
302          FROM dim_products
303          WHERE product_id = existing_product_id;
304
305      IF changes_needed THEN
306          --Expire the old version
307          UPDATE dim_products
308          SET current_flag = FALSE,
309              expiry_date = CURRENT_DATE
310          WHERE product_id = existing_product_id;
311
312          --Insert new version
313          INSERT INTO dim_products (
314              cdph_id, product_name, product_type, source_system, current_flag, entry_date, expiry_date
315          )
316          VALUES (
317              p_cdph_id, p_product_name, p_product_type, p_source_system, TRUE, CURRENT_DATE, '9999-12-31'
318          );
319      END IF;
320
321      ELSE
322          --No existing active version: insert new record
323          INSERT INTO dim_products (
324              cdph_id, product_name, product_type, source_system, current_flag, entry_date, expiry_date
325          )
326          VALUES (
327              p_cdph_id, p_product_name, p_product_type, p_source_system, TRUE, CURRENT_DATE, '9999-12-31'
328          );
329      END IF;
330  END;
331
332
333
334  --testing merge
335  CALL insert_dim_product_scd2_merge(
336      7151,
337      'Bluebell Shower Gel (New)', --new name
338      'Bath Products',
339      'COSMETICS'
340  );
341
342
343  select * from dim_products where cdph_id = 7151
```

Before Insert:

```
273 select * from dim_products where cdph_id = 7151
```

Data Output Messages Notifications								
Showing rows: 1 to 2 Page No: 1 of 1								
	product_id [PK] integer	cdph_id integer	product_name character varying (255)	product_type character varying (100)	source_system character varying (20)	current_flag boolean	entry_date date	expiry_date date
1	1	7151	Bluebell Bath & Shower Gel	Bath Products	COSMETICS	false	2025-04-25	2025-05-02
2	42198	7151	Bluebell Shower Gel	Bath Products	COSMETICS	true	2025-05-02	9999-12-31

After Insert:

Data Output Messages Notifications								
Showing rows: 1 to 3 Page No: 1 of 1								
	product_id [PK] integer	cdph_id integer	product_name character varying (255)	product_type character varying (100)	source_system character varying (20)	current_flag boolean	entry_date date	expiry_date date
1	1	7151	Bluebell Bath & Shower Gel	Bath Products	COSMETICS	false	2025-04-25	2025-05-02
2	42198	7151	Bluebell Shower Gel	Bath Products	COSMETICS	false	2025-05-02	2025-05-02
3	42199	7151	Bluebell Shower Gel (New)	Bath Products	COSMETICS	true	2025-05-02	9999-12-31

2. Using Tableau for Visualization

File Data Server Window Help

Connections: localhost PostgreSQL

Database: termProject

Table:

- dim_chemicals
- dim_company
- dim_cumulative_date
- dim_date
- dim_manufacturer_country
- dim_products
- dim_tested_by
- fact_chemical_in...etics_cumulative
- fact_chemical_in...metics_snapshot
- fact_metal_in...oducts_cumulative
- fact_metal_in...oducts_snapshot
- staging_consumer
- staging_cosmetics

fact_chemical_in_cosmetics_snapshot+ (termProje...

Connection: Live Extract

Filters: 0 Add

fact_chemical_in_cosme...

dim_chemicals

dim_company

dim_date

dim_products

fact_chemical_in_cosme...

Table Details:

#	#	#	#	#
fact_chemical_in_cosmetics_snapshot	fact_chemical_in_cosmetics_snapshot	fact_chemical_in_cosmetics_snapshot	fact_chemical_in_cosmetics_snapshot	fact_chemical_in_cosmetics_snapshot
Cosmetic Snapshot Fact Id	Product Id	Chemical Id	Company Id	Tester Id

Update Now

Update Automatically

Go to Worksheet