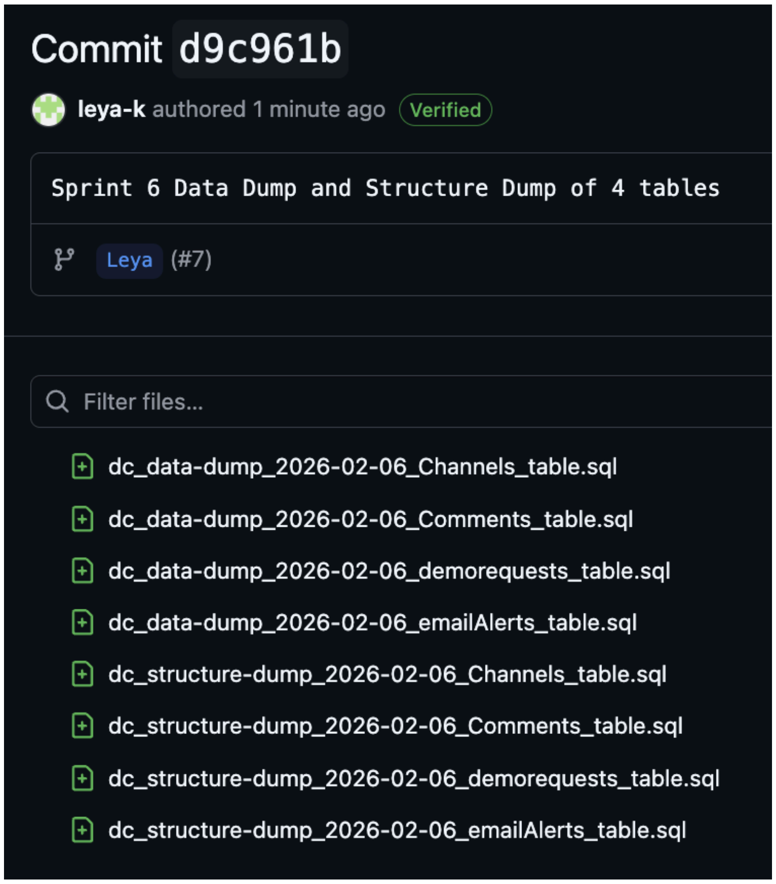
Task 1: Create a database dump and store it in GitHub



Commit d9c961b

leya-k authored 1 minute ago Verified

```
Sprint 6 Data Dump and Structure Dump of 4 tables
```

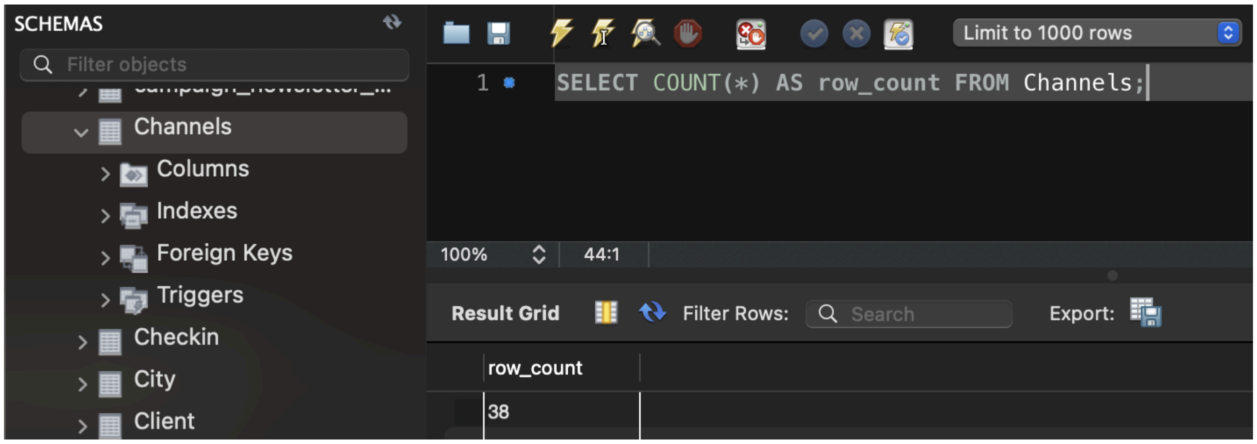
Leya (#7)

Filter files...

- dc_data-dump_2026-02-06_Channels_table.sql
- dc_data-dump_2026-02-06_Comments_table.sql
- dc_data-dump_2026-02-06_demorequests_table.sql
- dc_data-dump_2026-02-06_emailAlerts_table.sql
- dc_structure-dump_2026-02-06_Channels_table.sql
- dc_structure-dump_2026-02-06_Comments_table.sql
- dc_structure-dump_2026-02-06_demorequests_table.sql
- dc_structure-dump_2026-02-06_emailAlerts_table.sql

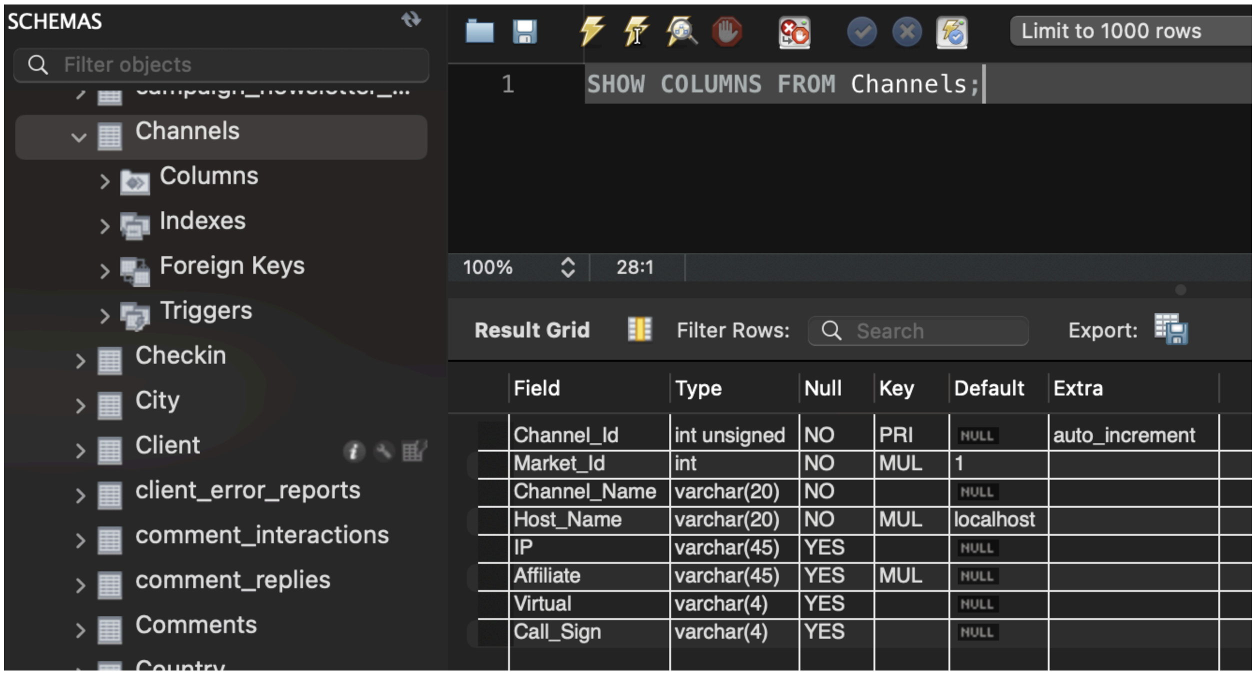
Task 2: Identify and classify large vs. small tables

We first performed a row and column count check for each table. This will show us the metadata of the table



```
SELECT COUNT(*) AS row_count FROM Channels;
```

row_count
38



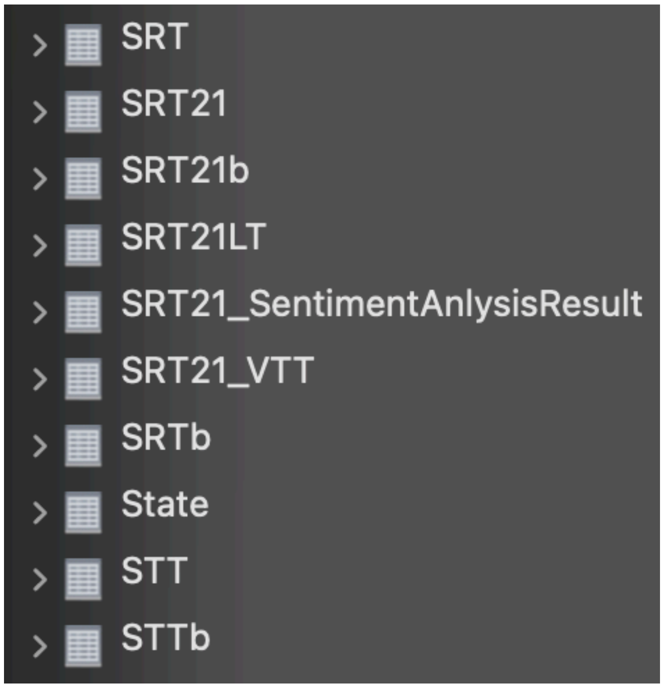
```
SHOW COLUMNS FROM Channels;
```

Field	Type	Null	Key	Default	Extra
Channel_Id	int unsigned	NO	PRI	NULL	auto_increment
Market_Id	int	NO	MUL	1	
Channel_Name	varchar(20)	NO		NULL	
Host_Name	varchar(20)	NO	MUL	localhost	
IP	varchar(45)	YES		NULL	
Affiliate	varchar(45)	YES	MUL	NULL	
Virtual	varchar(4)	YES		NULL	
Call_Sign	varchar(4)	YES		NULL	

Results: All tables are small tables. This means that they are all suitable for a data dump.

Task 3: Validate SRT table data retention window

In the database, the SRT tables are labeled as SRT in the beginning. Therefore, the 4 assigned tables are not SRT tables. This is how SRT tables are shown:

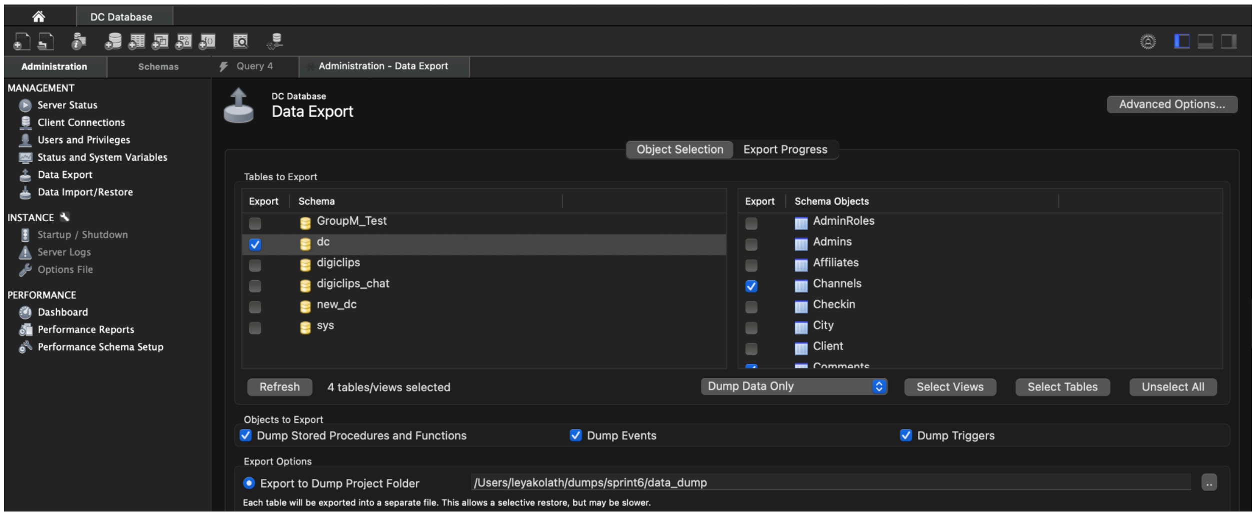


SRT
SRT21
SRT21b
SRT21LT
SRT21_SentimentAnlysisResult
SRT21_VTT
SRTb
State
STT
STTb

These are how the SRT tables are labelled in the database.

Task 4: Split the database dump into two logical dumps

This is the database administrator. From this portal we had opened the dc database and selected the 4 tables: Channels, Comments, Demo_requests, and emailAlerts.



DC Database
Data Export

Tables to Export

Export	Schema
☐	GroupM_Test
✓	dc
☐	digiclips
☐	digiclips_chat
☐	new_dc
☐	sys

Export	Schema Objects
☐	AdminRoles
☐	Admins
☐	Affiliates
✓	Channels
☐	Checkin
☐	City
☐	Client
✓	Comments

4 tables/views selected

Dump Data Only

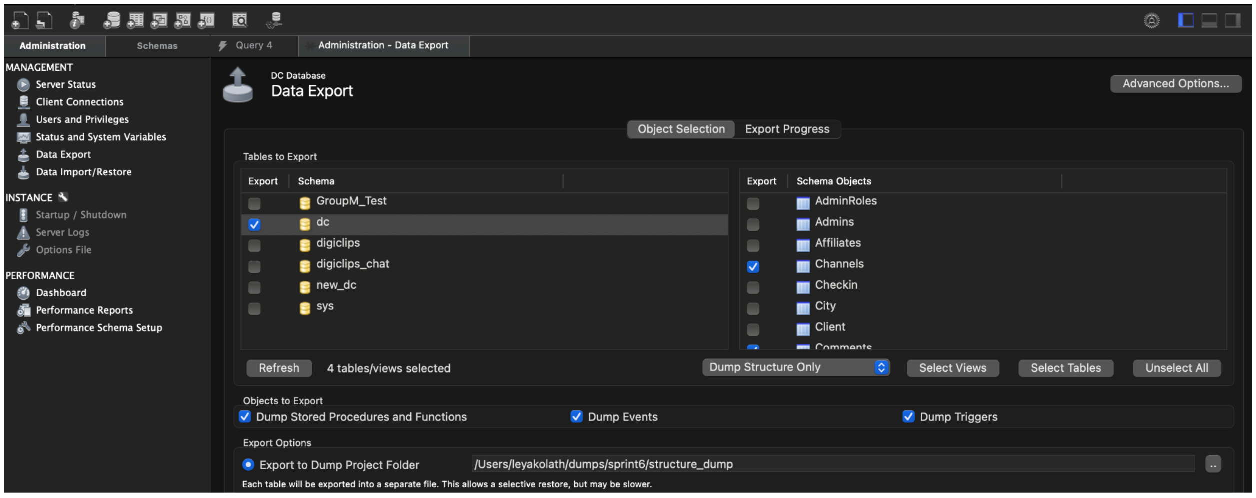
Objects to Export

✓ Dump Stored Procedures and Functions ✓ Dump Events ✓ Dump Triggers

Export Options

● Export to Dump Project Folder /Users/leyakolath/dumps/sprint6/data_dump

Each table will be exported into a separate file. This allows a selective restore, but may be slower.



DC Database
Data Export

Tables to Export

Export	Schema
☐	GroupM_Test
✓	dc
☐	digiclips
☐	digiclips_chat
☐	new_dc
☐	sys

Export	Schema Objects
☐	AdminRoles
☐	Admins
☐	Affiliates
✓	Channels
☐	Checkin
☐	City
☐	Client
✓	Comments

4 tables/views selected

Dump Structure Only

Objects to Export

✓ Dump Stored Procedures and Functions ✓ Dump Events ✓ Dump Triggers

Export Options

● Export to Dump Project Folder /Users/leyakolath/dumps/sprint6/structure_dump

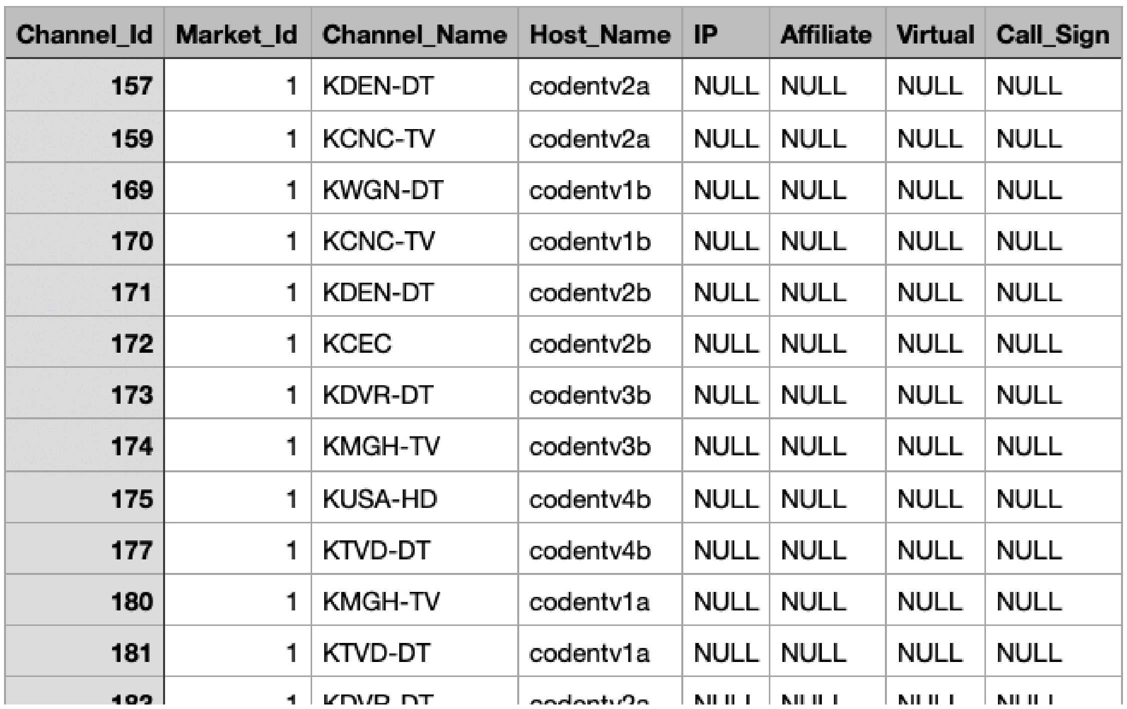
Each table will be exported into a separate file. This allows a selective restore, but may be slower.

We performed Dump A. We separated the dump into a data dump and a structured dump of the tables.

Task 5: Extract and download data that should not be dumped

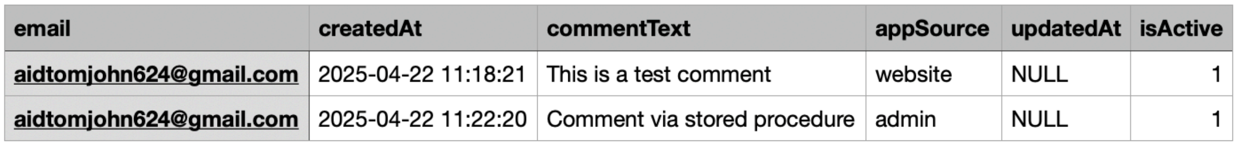
We extracted and downloaded the data as csv files. The table demo_requests was empty.

Channels_data



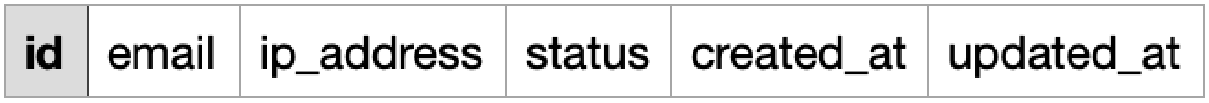
Channel_Id	Market_Id	Channel_Name	Host_Name	IP	Affiliate	Virtual	Call_Sign
157	1	KDEN-DT	codentv2a	NULL	NULL	NULL	NULL
159	1	KCNC-TV	codentv2a	NULL	NULL	NULL	NULL
169	1	KWGN-DT	codentv1b	NULL	NULL	NULL	NULL
170	1	KCNC-TV	codentv1b	NULL	NULL	NULL	NULL
171	1	KDEN-DT	codentv2b	NULL	NULL	NULL	NULL
172	1	KCEC	codentv2b	NULL	NULL	NULL	NULL
173	1	KDVR-DT	codentv3b	NULL	NULL	NULL	NULL
174	1	KMGH-TV	codentv3b	NULL	NULL	NULL	NULL
175	1	KUSA-HD	codentv4b	NULL	NULL	NULL	NULL
177	1	KTVD-DT	codentv4b	NULL	NULL	NULL	NULL
180	1	KMGH-TV	codentv1a	NULL	NULL	NULL	NULL
181	1	KTVD-DT	codentv1a	NULL	NULL	NULL	NULL
183	1	KDVR-DT	codentv2a	NULL	NULL	NULL	NULL

Comments_data



email	createdAt	commentText	appSource	updatedAt	isActive
aidtomjohn624@gmail.com	2025-04-22 11:18:21	This is a test comment	website	NULL	1
aidtomjohn624@gmail.com	2025-04-22 11:22:20	Comment via stored procedure	admin	NULL	1

demo_requests



id	email	ip_address	status	created_at	updated_at

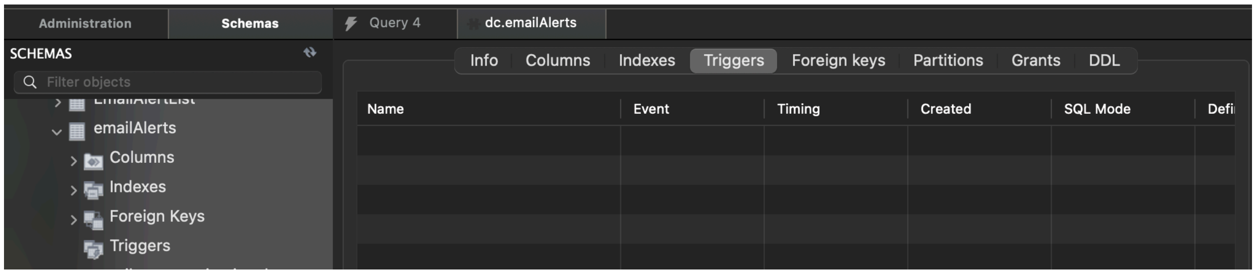
emailAlerts has 39 columns so the screenshot was too big to include.

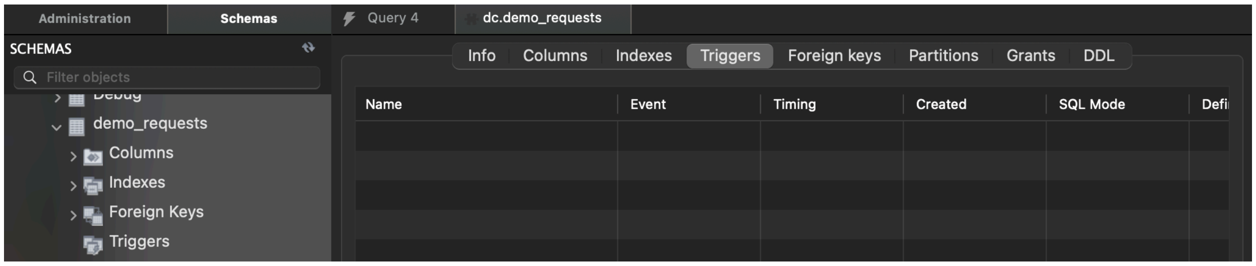
Task 6: Escalate retention violations if found

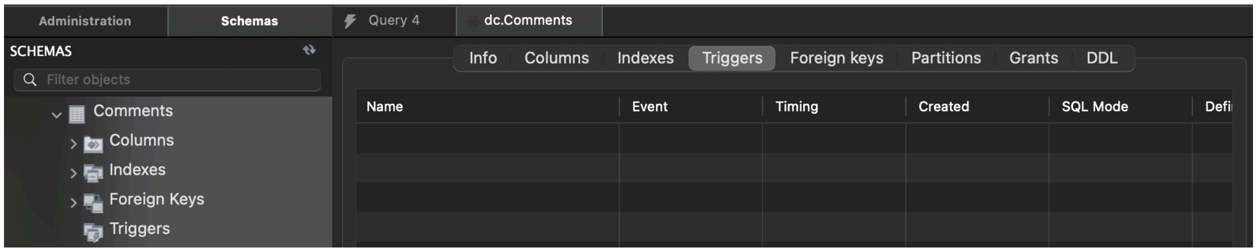
The Comments and emailAlerts tables all have inserted data older than 90 days. We will reach out to the team to get that data cleared out. Once it is removed, we will provide new entries into the database

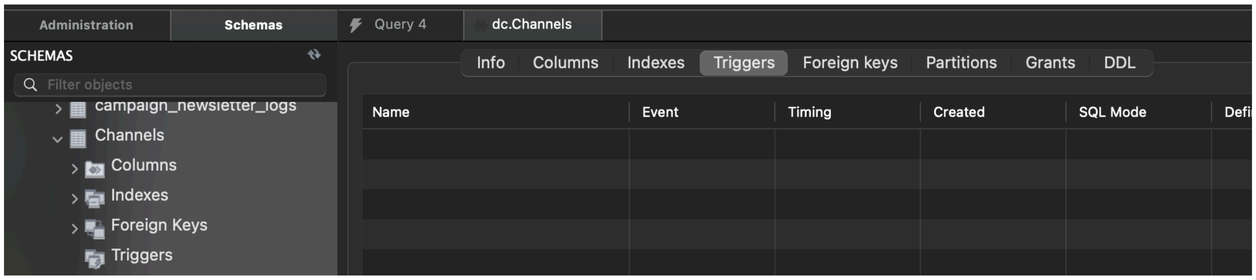
** Task 7: Trigger Check

None of the assigned tables contain triggers.









Scenario 1: Confirm Table Qualifies to be Dumped

Description: Check that the tables contain data that does not need to be kept (e.g., debug data, large schemas). Data that must be preserved (e.g., emails, names) should be downloaded to GitHub instead of dumped.

Task Completed: We preserved the table data by downloading them as csv files before performing the dump.

Scenario 2: Verify Tables to be Dumped

Description: Before performing the dump, verify the list of tables that are selected to be dumped with DigiClips to ensure that no important information is lost. Once the dump is executed, there is no retrieving the lost data.

Task Completed: We confirmed with DigiClips that these tables were fine to dump, as they are not necessary to preserve in their current state.

Scenario 3: Verify Tables to be Downloaded

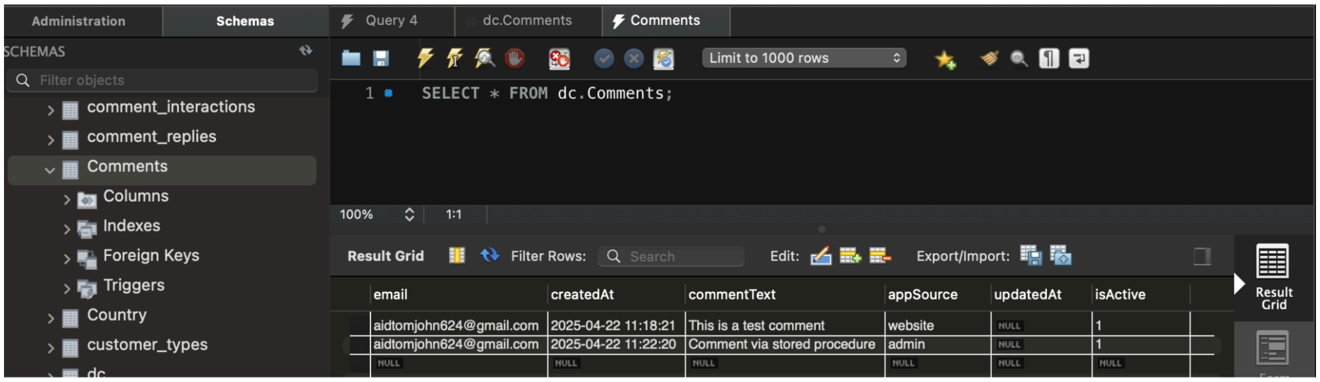
Description: Before performing the dump, verify the list of tables that are selected to be downloaded to GitHub with DigiClips to ensure that all the kept tables contain data that is necessary to maintain. Large downloaded tables that contain redundant data will slow down the efficiency of the database.

Task Completed: None of the tables were SRT tables so we didn't have to worry about SRT retention. We downloaded the data as csv files for retention safety.

Scenario 4: Verify Age of Web Scraped Data

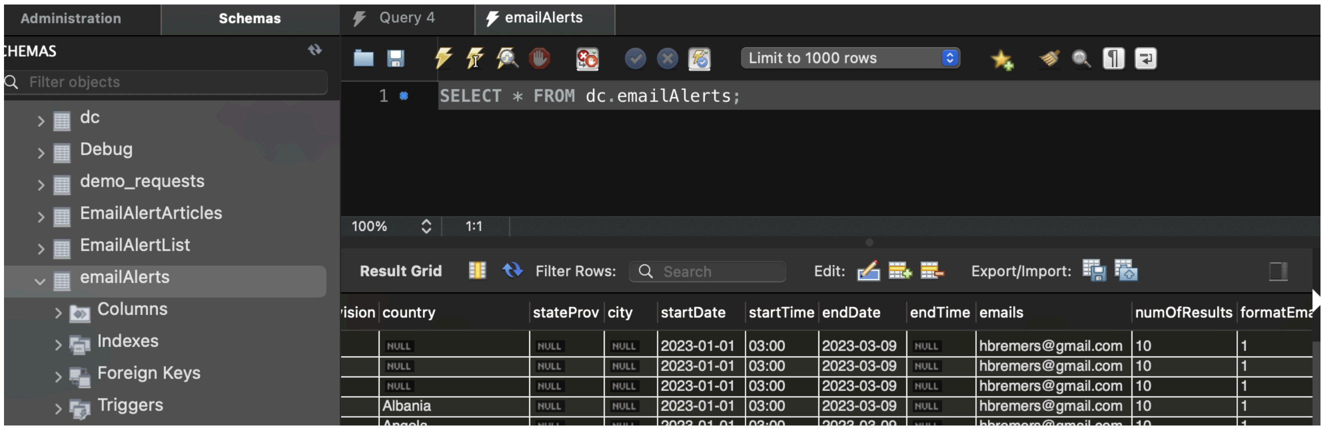
Description: Inspect data within each table to ensure that any web scraping records contained are no older than 90 days. If the age of the data exceeds this threshold, it should be dumped to maintain accurate, up to date records. The backend team or DigiClips should be notified if such records exist, so as to promote regular maintenance of the database.

Task Completed: The Comments and emailAlerts table has data older than 90 days. We will notify the team of this and have them scrub the data.



```
SELECT * FROM dc.Comments;
```

email	createdAt	commentText	appSource	updatedAt	isActive
aidtomjohn624@gmail.com	2025-04-22 11:18:21	This is a test comment	website	NULL	1
aidtomjohn624@gmail.com	2025-04-22 11:22:20	Comment via stored procedure	admin	NULL	1
NULL	NULL	NULL	NULL	NULL	NULL



```
SELECT * FROM dc.emailAlerts;
```

ision	country	stateProv	city	startDate	startTime	endDate	endTime	emails	numOfResults	formatEma
	NULL	NULL	NULL	2023-01-01	03:00	2023-03-09	NULL	hbremers@gmail.com	10	1
	NULL	NULL	NULL	2023-01-01	03:00	2023-03-09	NULL	hbremers@gmail.com	10	1
	NULL	NULL	NULL	2023-01-01	03:00	2023-03-09	NULL	hbremers@gmail.com	10	1
	Albania	NULL	NULL	2023-01-01	03:00	2023-03-09	NULL	hbremers@gmail.com	10	1