

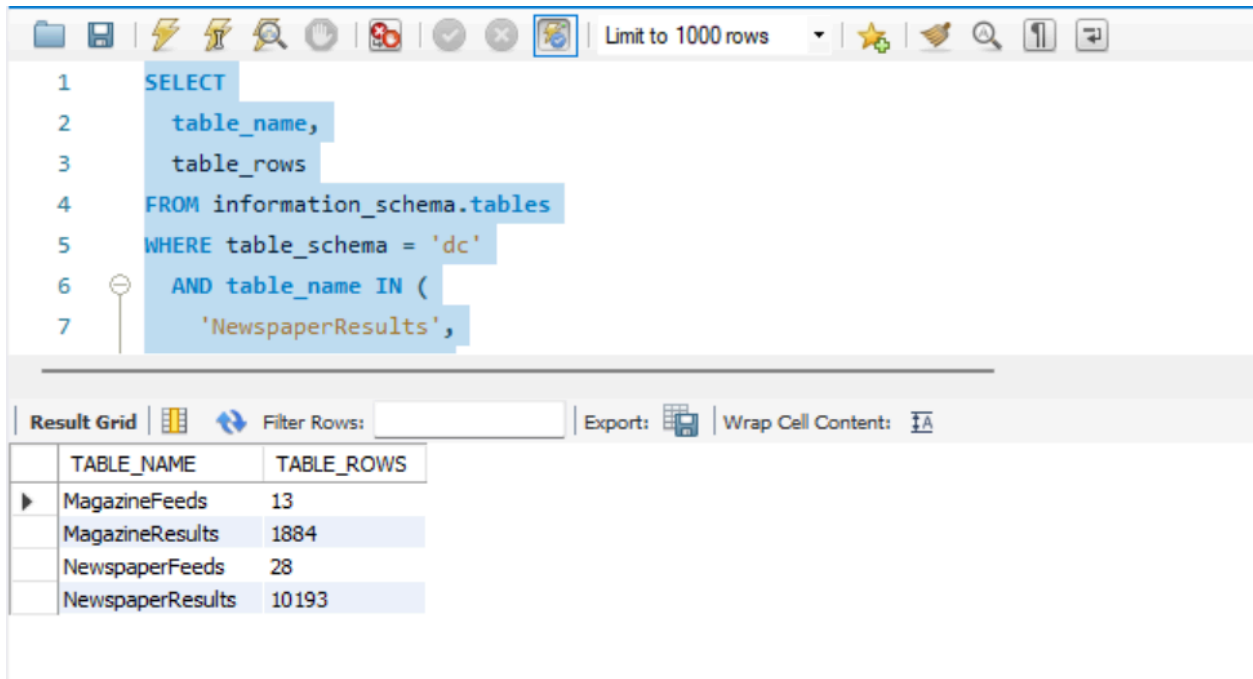
Reporting Tables Implementation

Sprint 6: NewspaperFeeds, MagazineFeeds, NewspaperResults and MagazineResults

Overview

This report explains data preservation methods used for the assigned tables.

Dumps were created in two categories based on table size, volatility, and retention rules.



The screenshot shows a SQL query editor with the following query:

```
1 SELECT
2   table_name,
3   table_rows
4 FROM information_schema.tables
5 WHERE table_schema = 'dc'
6    AND table_name IN (
7       'NewspaperResults',
```

Below the query editor, the 'Result Grid' is displayed, showing the results of the query:

TABLE_NAME	TABLE_ROWS
MagazineFeeds	13
MagazineResults	1884
NewspaperFeeds	28
NewspaperResults	10193

Table	Size	Dump Type	Notes
NewspaperFeeds	Small	Full Dump (A)	Includes structure + all data and manipulations
MagazineFeeds	Small	Full Dump (A)	Includes structure + all data and manipulations

NewspaperResults	Large, scraped continuously	Structure only (B)	Data excluded due to size and retention; schema preserved for rollback/testing
MagazineResults	Large, scraped continuously	Structure Only (B)	Data excluded due to size and retention; schema preserved for rollback/testing

Retention check: All tables were checked for records older than 90 days. Any violations were flagged and will be escalated to the backend teams and sponsors.

Dump Details

Dump A → Small Tables

- Tables included: NewspaperFeeds, MagazineFeeds
- Export type: Structure + Data
- File names:
 - dc_structure-dump_2026-02-07_MagazineFeeds.sql
 - dc_data-dump_2026-02-07_MagazineFeeds.sql
 - dc_structure-dump_2026-02-07_NewspaperFeeds.sql
 - dc_data-dump_2026-02-07_NewspaperFeeds.sql
- Purpose: Provides full data for rollback, testing, and validation for smaller, stable tables.

Dump B → LargeTables

- Tables included: NewspaperResults, MagazineResults
- Export type: Structure only (no data)
- File name:
 - dc_structure-dump_2026-02-08_MagazineResults.sql
 - dc_structure-dump_2026-02-08_NewspaperResults.sql
- Purpose: Preserves table schema and indexes for large, volatile scraping tables without storing millions of rows of temporary data.

Notes:

Triggers, events, and stored procedures were excluded from the schema-only dumps because they are not required for rollback or performance testing.

Refer to sprint 1 and sprint 2 reporting optimization documentation for a detailed explanation of table schemas, stored procedures and events.

NewspaperResults & MagazineResults (Large Tables)

Oldest and Newest Records of NewspaperResults:

```
1  SELECT
2      MIN(AddedDate) AS oldest_record,
3      MAX(AddedDate) AS newest_record
4  FROM NewspaperResults;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

	oldest_record	newest_record
▶	2026-01-13 16:51:05	2026-02-09 09:27:38

Oldest and Newest Records of MagazineResults:

| Limit to 1000 rows

```
1  SELECT
2      MIN(AddedDate) AS oldest_record,
3      MAX(AddedDate) AS newest_record
4  FROM MagazineResults;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

	oldest_record	newest_record
▶	2026-01-13 16:51:03	2026-02-09 09:27:30

- There is a cleanup procedure that runs as an event for both of these tables. These tables mainly serve as a cache for scraped data.

Download / Excluded Data

Some data was excluded from dumps due to size and retention rules.

These are the steps followed to identify such tables:

1. Identified data that cannot be included in SQL dumps
 - a. Historical scrape data from over 90 days
 - b. Large result sets that would create huge files
2. Downloaded and exported data separately”
 - a. Format: csv file
3. Stored securely in:
 - a. /dumps folder in Github repository
 - b. Explicitly named:
 - i. dc_data-download_2026-02-09_MagazineResults.csv
 - ii. dc_data-download_2026-02-09_NewspaperResults.csv
 - c. Ensured it is accessible for future validation

MagazineResults csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	Title	Link	Author	Summary	Magazine	PublishDate	MagazineL	ImageTitle	ImageURL	UpdateDate	Categories	GUID	AddedDate
2	542254	The 51 Bes	https://ww	Jennifer M. A	Thousan	WIRED	13/01/2026 13:00	https://ww	NULL	https://me	13/01/2026 13:00	Culture,Ci	63d96ba4	13/01/2026 16:51
3	542255	Everything	https://ww	Vittoria El	In the secc	WIRED	13/01/2026 11:59	https://ww	NULL	https://me	13/01/2026 11:59	Politics,Ci	69603ce6	13/01/2026 16:51
4	542256	Robloxâ€	https://ww	David Gilb	Kids are b	WIRED	13/01/2026 11:54	https://ww	NULL	https://me	13/01/2026 11:54	Culture,Ci	696130b5	13/01/2026 16:51
5	542257	Shortcut Y	https://ww	Brad Bourc	Elgatoâ€	WIRED	13/01/2026 11:44	https://ww	NULL	https://me	13/01/2026 11:44	Gear,Gear	6966868e	13/01/2026 16:51
6	542258	Chinaâ€	https://ww	Zeyi Yang	Are You D	WIRED	13/01/2026 11:27	https://ww	NULL	https://me	13/01/2026 11:27	Business,f	69667166	13/01/2026 16:51
7	542259	Dozens of	https://ww	Caroline H	A contract	WIRED	13/01/2026 09:45	https://ww	NULL	https://me	13/01/2026 09:45	Security,Si	69665b09	13/01/2026 16:51
8	542260	Microsoft	https://ww	Molly Taft	In respons	WIRED	13/01/2026 07:48	https://ww	NULL	https://me	13/01/2026 07:48	Science,Ti	69664e3b	13/01/2026 16:51
9	542261	Urevo Spa	https://ww	Michael V	Urevoâ€	WIRED	13/01/2026 06:00	https://ww	NULL	https://me	13/01/2026 06:00	Gear,Gear	695ece45	13/01/2026 16:51
10	542262	Board Revi	https://ww	Simon Hill	I loved Boa	WIRED	13/01/2026 05:00	https://ww	NULL	https://me	13/01/2026 05:00	Gear,Gear	695d2097	13/01/2026 16:51
11	542263	Atonemo S	https://ww	Chris Hasl	Can this d	WIRED	13/01/2026 04:30	https://ww	NULL	https://me	13/01/2026 04:30	Gear,Gear	69652aad	13/01/2026 16:51
12	542264	Reid Hoffn	https://ww	Katie Drun	The Linked	WIRED	13/01/2026 04:30	https://ww	NULL	https://me	13/01/2026 04:30	Business,f	696181d0	13/01/2026 16:51
13	542265	Trump Dec	https://ww	Noah Shac	If you want	WIRED	13/01/2026 04:00	https://ww	NULL	https://me	13/01/2026 04:00	The Big Stc	691cca2c	13/01/2026 16:51
14	542266	What to D	https://ww	Lily Hay N	eWith feder	WIRED	13/01/2026 03:30	https://ww	NULL	https://me	13/01/2026 03:30	Security,Si	695fc7a4	13/01/2026 16:51

NewspaperResults csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	Title	Link	Author	Summary	Newspape	PublishDate	NewspaperLink	ImageTitle	ImageURL	UpdateDate	Categories	GUID	AddedDate
2	6045654	Education,	http://www	NULL	China to l	China Dail	11/12/2017 17:59	http://usa.chinadaily.con	NULL	NULL	11/12/2017 17:59	Education	http://www	13/01/2026 16:51
3	6045656	Internatio	http://www	NULL	High-leve	China Dail	11/12/2017 17:46	http://usa.chinadaily.con	NULL	NULL	11/12/2017 17:46	Environme	http://www	13/01/2026 16:51
4	6045659	Beijing, Mc	http://www	NULL	China and	China Dail	11/12/2017 17:41	http://usa.chinadaily.con	NULL	NULL	11/12/2017 17:41	Military	http://www	13/01/2026 16:51
5	6045660	Memorial l	http://www	NULL	A memori	China Dail	11/12/2017 17:39	http://usa.chinadaily.con	NULL	NULL	11/12/2017 17:39	Society	http://www	13/01/2026 16:51
6	6045662	More milit	http://www	NULL	Greater ir	China Dail	11/12/2017 17:38	http://usa.chinadaily.con	NULL	NULL	11/12/2017 17:38	Military	http://www	13/01/2026 16:51
7	6045666	Dangerous	http://www	NULL	Multiple s	China Dail	11/12/2017 17:36	http://usa.chinadaily.con	NULL	NULL	11/12/2017 17:36	Society	http://www	13/01/2026 16:51
8	6045668	Taiwan's N	http://www	NULL	Delegatio	China Dail	11/12/2017 17:33	http://usa.chinadaily.con	NULL	NULL	11/12/2017 17:33	HK Macao	http://www	13/01/2026 16:51
9	6045671	Former Ch	http://www	NULL	Sun Zheng	which may	bail	residential surveillance	detention	China Dail	11/12/2017 17:28	http://usa.	NULL	NULL
10	6045675	Satellite lo	http://www	NULL	 	China Dail	11/12/2017 17:27	http://usa.chinadaily.con	NULL	NULL	11/12/2017 17:27	Innovation	http://www	13/01/2026 16:51
11	6045680	Water dive	http://www	NULL	Resider	China Dail	11/12/2017 16:57	http://usa.chinadaily.con	NULL	NULL	11/12/2017 16:57	Society	http://www	13/01/2026 16:51
12	6045682	Oldest Nai	http://www	NULL	Guan	China Dail	11/12/2017 16:32	http://usa.chinadaily.con	NULL	NULL	11/12/2017 16:32	Newsmake	http://www	13/01/2026 16:51
13	6045685	AllIB loans	http://www	NULL	A worke	China Dail	11/12/2017 16:27	http://usa.chinadaily.con	NULL	NULL	11/12/2017 16:27	Society	http://www	13/01/2026 16:51
14	6045688	Discoverin	http://www	NULL	 	China Dail	11/12/2017 16:09	http://usa.chinadaily.con	NULL	NULL	11/12/2017 16:09	Featured C	http://www	13/01/2026 16:51
15	6045691	Xi: Uproot	http://www	NULL	Fight to c	China Dail	11/12/2017 14:51	http://usa.chinadaily.con	NULL	NULL	11/12/2017 14:51	National a	http://www	13/01/2026 16:51

Notes:

Full historical scrape data cannot be included due to scheduled cleanups but a small current subset has been downloaded.

Trigger Checks

None of my assigned tables have triggers.

SRT Validation

None of the tables assigned to me this sprint are SRT tables.

Next Steps

- Verify that dumps restore correctly in test environments.
- Document any retention violations and escalations
- Track dumps and documentation from other teams
- Make optimization changes to the dumped tables

Summary

- Large, volatile tables (Results) → schema-only dumps
- Small, stable tables (Feeds) → full dumps
- Excluded/large data → downloaded separately, stored securely
- Documentation ensures traceability, rollback support, and compliance with retention rules