

Bing News Streaming Project

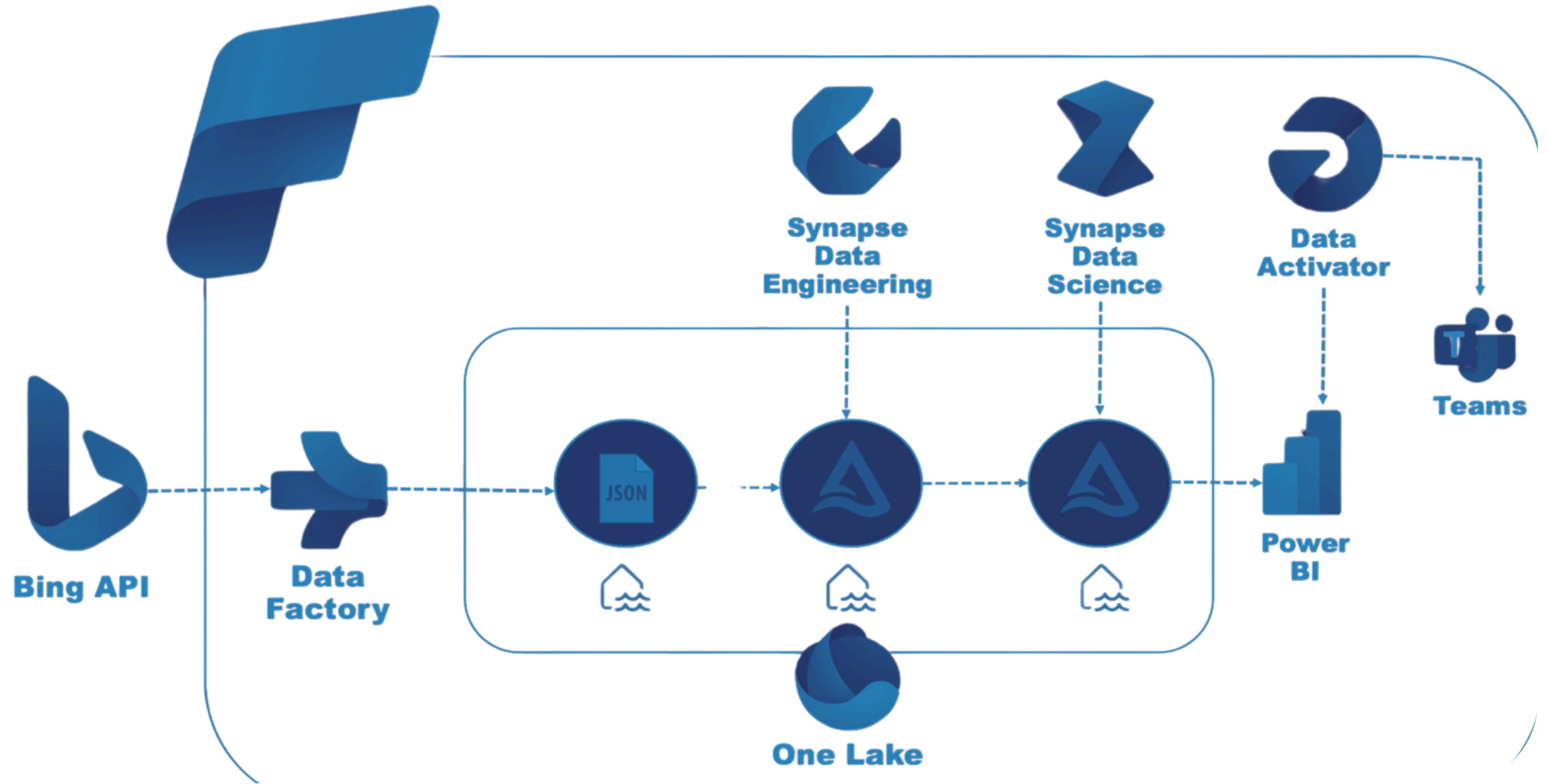
using Microsoft Fabric



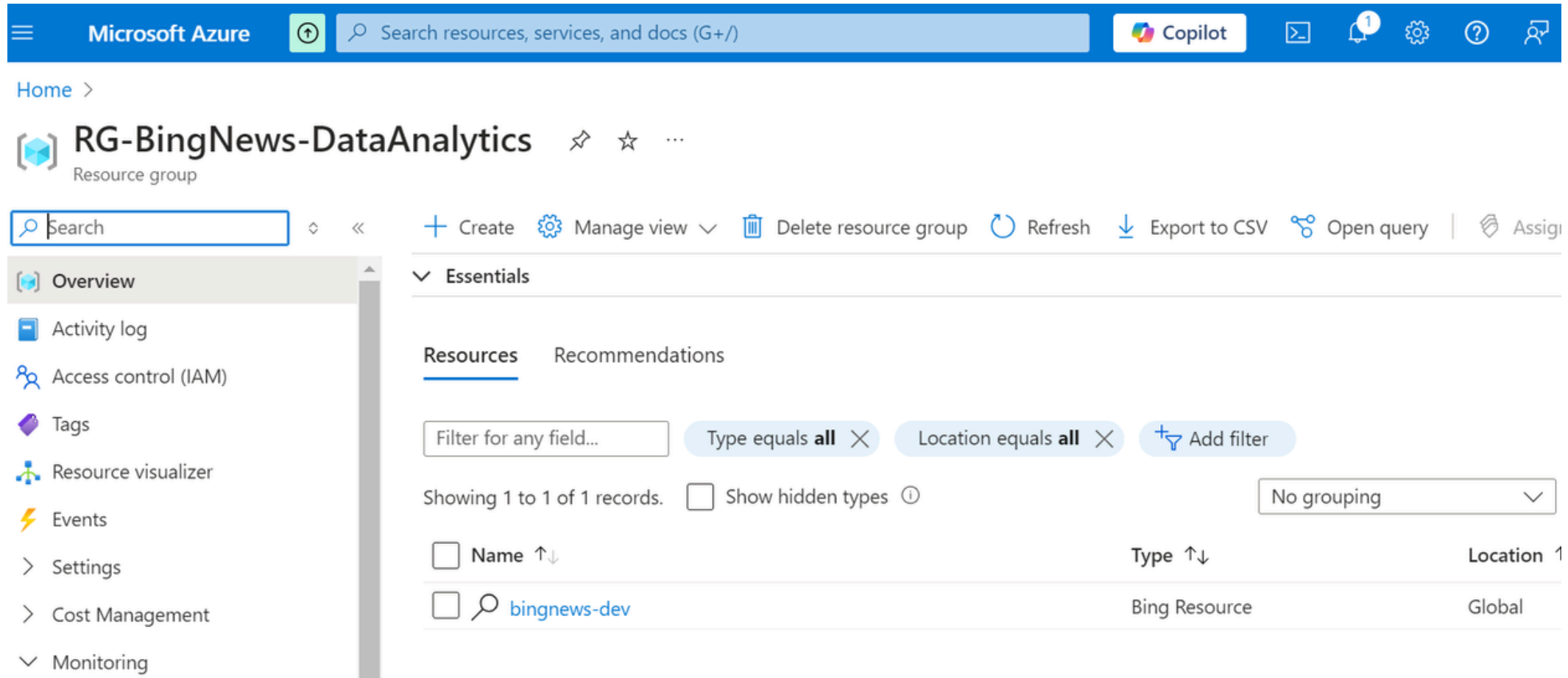
Objective

To get last 24 hrs news published on BING and create a dashboard to display 'Positive' , 'Neutral' and 'Negative' news

Architecture



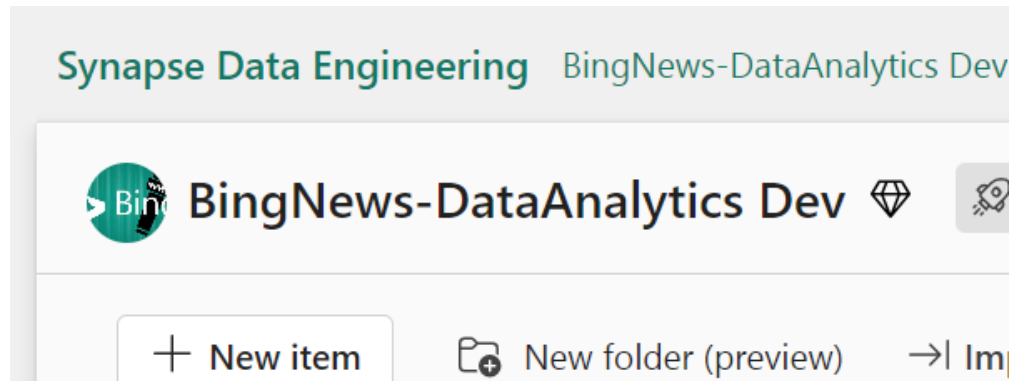
1. Setup Resource group and add a resource for Bing Search API



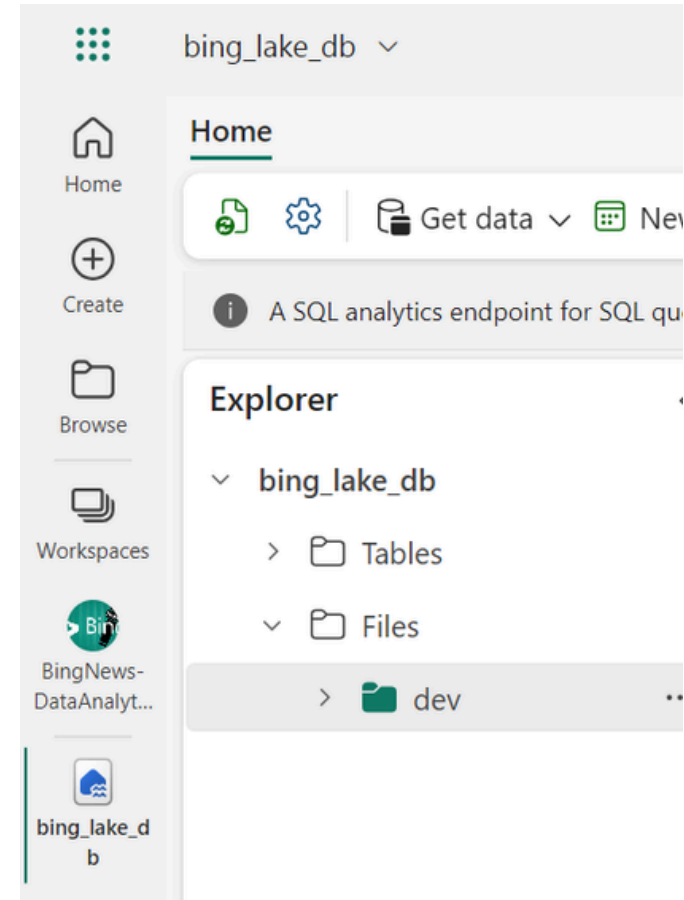
The screenshot shows the Microsoft Azure portal interface. At the top, the header includes the Microsoft Azure logo, a search bar, and various utility icons. Below the header, the breadcrumb 'Home >' is followed by the resource group name 'RG-BingNews-DataAnalytics'. A left-hand navigation pane lists various management tools like Activity log, Access control (IAM), Tags, Resource visualizer, Events, Settings, Cost Management, and Monitoring. The main content area is titled 'Essentials' and contains a 'Resources' tab. Below this tab, there are filter buttons for 'Type equals all' and 'Location equals all', along with a search filter box. A status line indicates 'Showing 1 to 1 of 1 records.' and a 'Show hidden types' checkbox. A table lists the resources, with one entry: 'bingnews-dev' of type 'Bing Resource' in the 'Global' location.

Name	Type	Location
bingnews-dev	Bing Resource	Global

2. Create Dev workspace in Fabric Synapse Data Engineering

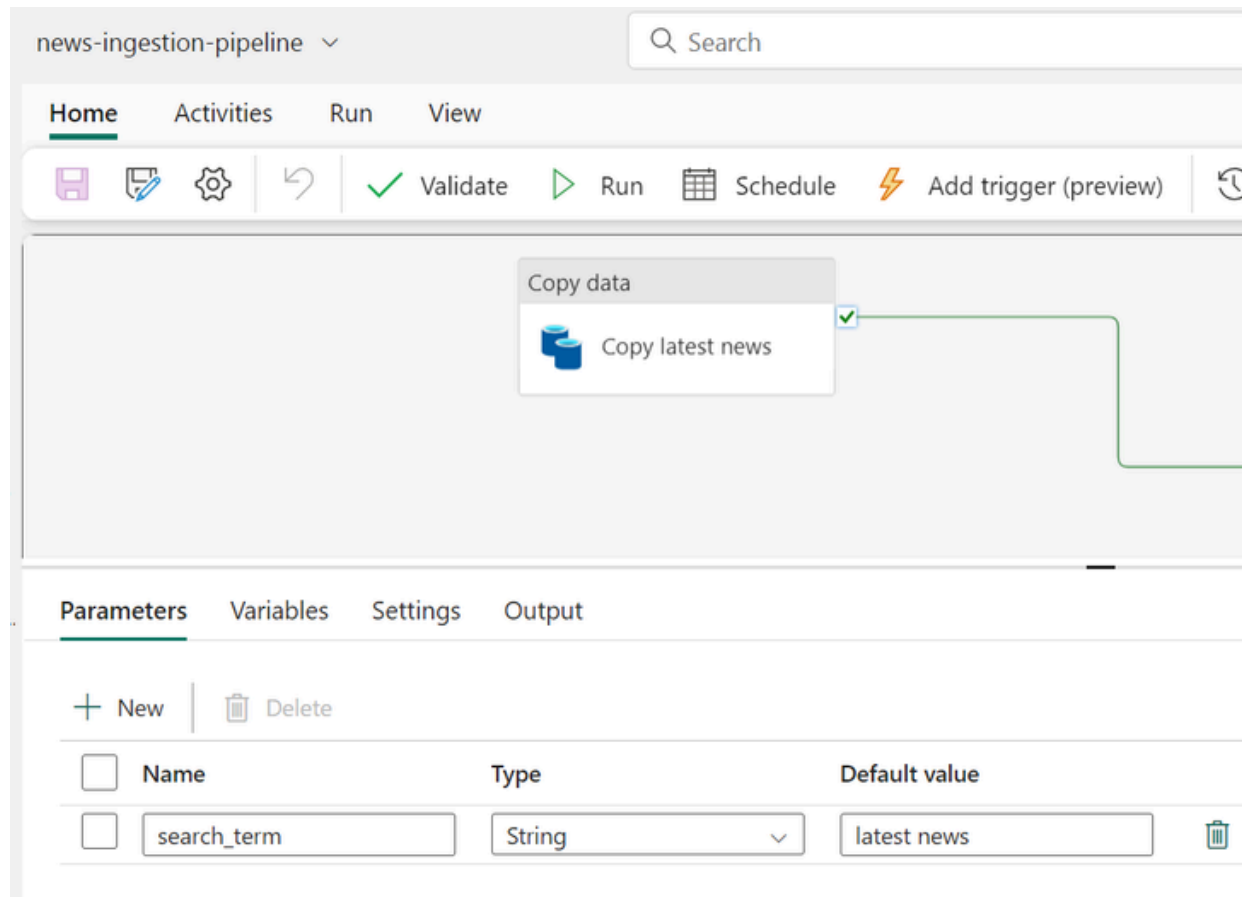


3. Create a lakehouse inside the workspace, create a dev folder inside Files storage



4. Go to Data Factory and Create a New Pipeline

Add parameter 'search_term'.

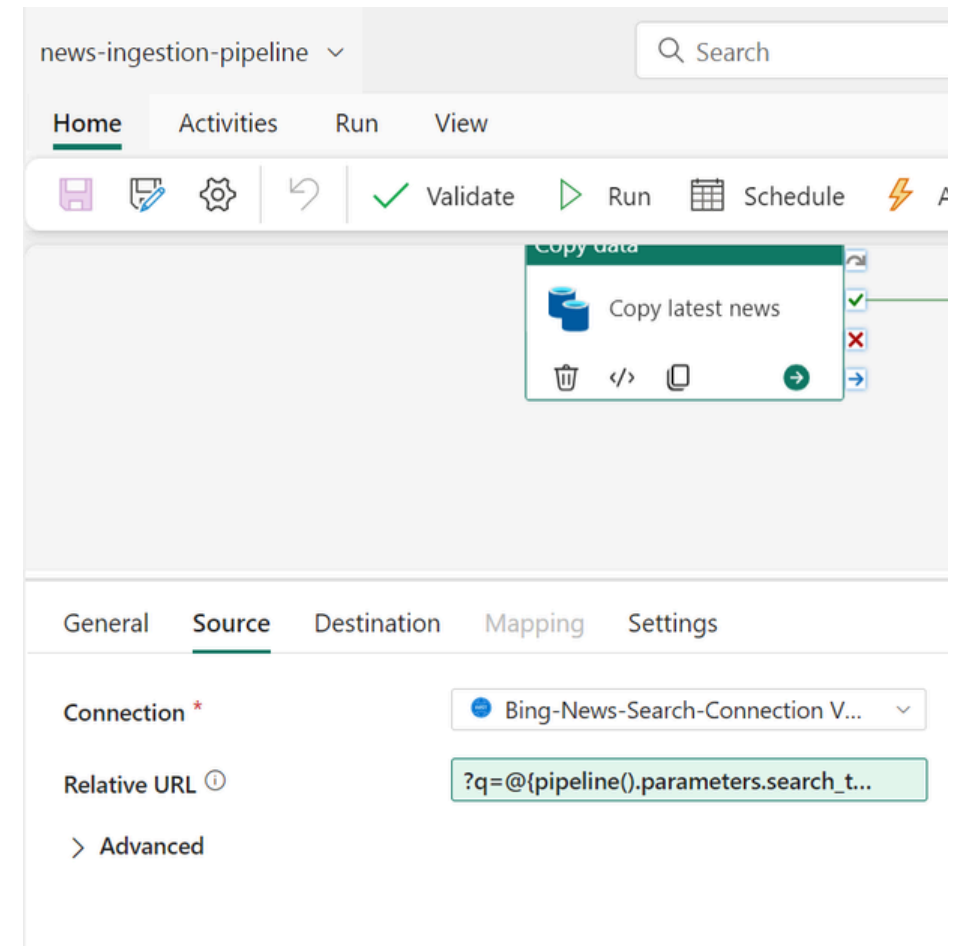


The screenshot shows the Azure Data Factory pipeline editor for a pipeline named 'news-ingestion-pipeline'. The 'Activities' tab is selected, displaying a 'Copy data' activity named 'Copy latest news'. Below the canvas, the 'Parameters' tab is active, showing a table with one parameter: 'search_term' of type 'String' with a default value of 'latest news'.

Name	Type	Default value
search_term	String	latest news

5. Add a 'Copy' activity

Create a new Connection using REST API source and connecting it to the Bing Search API resource in Azure Portal. In destination using the Dev folder in Files in lakehouse. Run pipeline. File should be saved in json format.



The screenshot shows the configuration for the 'Copy latest news' activity in the 'Source' tab. The 'Connection' is set to 'Bing-News-Search-Connection V...'. The 'Relative URL' is set to '?q=@{pipeline().parameters.search_t...'. The 'Advanced' section is expanded.

Connection * Bing-News-Search-Connection V...

Relative URL ⓘ ?q=@{pipeline().parameters.search_t...

> Advanced

6. Go to Synapse Data Engineering and create a new notebook

- a. Read the JSON file as a Dataframe
- b. Do the processing of the raw data

```
1 from pyspark.sql.functions import explode
2 import json
3 df_exploded = df.select(explode(df["value"]).alias("json_object"))
4 display(df_exploded)
```

✓ - Command executed in 1 sec 570 ms by Mayuri Saikia on 1:15:40 AM, 10/07/24

Table | Chart | Download | Showing rows 1 - 51 | Inspector

	ANY json_object
1	{"datePublished":"2024-10-06T07:23:00.0000000Z","name":"Mumbai New..
2	{"datePublished":"2024-10-05T16:41:00.0000000Z","name":"Here are the l..
3	{"datePublished":"2024-10-06T01:47:00.0000000Z","name":"Israel-Iran Wa..
4	{"datePublished":"2024-10-06T14:01:54.0000000Z","name":"New report fi..

ANY json_object

Missing:

Unique:

- c. Write the final cleaned dataframe as a delta table into the Lakehouse as 'tbl_latest_news'. Add incremental loader code in SCD type 1. Run the notebook.

+ Warehouses

bing_lake_db

Schemas

dbo

Tables

Data preview - tbl_latest_news

Showing 1000 rows

Search

	ABC title	ABC description	ABC category	ABC url	ABC image	ABC provider	ABC datePublished
1	Kamala Harris over...	Kamala Harris has ...	Politics	https://www.msn.c...	https://www.bing....	WION on MSN.com	10-Oct-2024
2	Lithuanians to pick...	Lithuanians are ex...	World	https://www.reuter...	https://www.bing....	Reuters	10-Oct-2024
3	GM offering new h...	General Motors sa...	ScienceAndTechno...	https://www.reuter...	https://www.bing....	Reuters	10-Oct-2024
4	U.S. inflation conti...	Inflation in Septe...	Business	https://www.msn.c...	https://www.bing....	CBS News on MSN...	10-Oct-2024

7. Create a new notebook for news sentiment analysis

a. Import ML model AnalyzeText for text analysis and apply to the dataframe of the table for processed news data.

```
[3] 1 import synapse.ml.core
    2 from synapse.ml.services import AnalyzeText
```

✓ - Command executed in 3 sec 612 ms by Mayuri Saikia on 1:24:37 AM, 10/07/24

```
[4] 1 # Import the model and configure the input and output columns
    2
    3 model = (AnalyzeText()
    4         .setTextCol("description")
    5         .setKind("SentimentAnalysis")
    6         .setOutputCol("newsSentiment")
    7         .setErrorCol("error")
    8         )
```

✓ - Command executed in 845 ms by Mayuri Saikia on 1:29:44 AM, 10/07/24

```
[6] 1 #Apply the model to our dataframe
    2
    3 result = model.transform(df)
```










✓ - Command executed in 807 ms by Mayuri Saikia on 1:29:53 AM, 10/07/24

b. Do the required transformation to save the sentiment value in a column and save the dataframe in SCD type 1 incremental load format to the Lakehouse in a new table.


8. Run the pipeline after adding the notebooks

news-ingestion-pipeline ▾


Home Activities Run View

     Validate  Run  Schedule  Add trigger (preview)  View run h


Copy data

-  Copy latest news ✓



Notebook


-  Transform Raw Data to delta table ✓

Notebook

-  Perform Sentiment Analysis ✓

Parameters Variables Settings Output

 New  Delete

<input type="checkbox"/>	Name	Type	Default value	
<input type="checkbox"/>	search_term	String ▾	latest news	

9. Create new semantic data model and then create Power BI report

News Sentiment From Bing
PAGE 1

Insights summary
tbl_sentiment_analysed

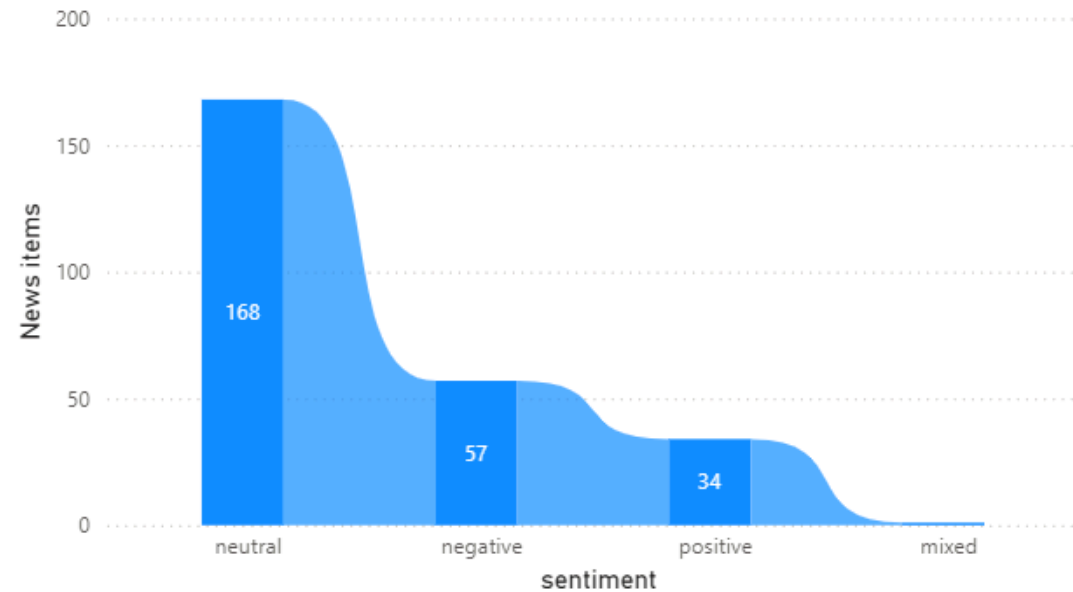
260
Count of title

261
Count of url

datePublished

All

News items by sentiment

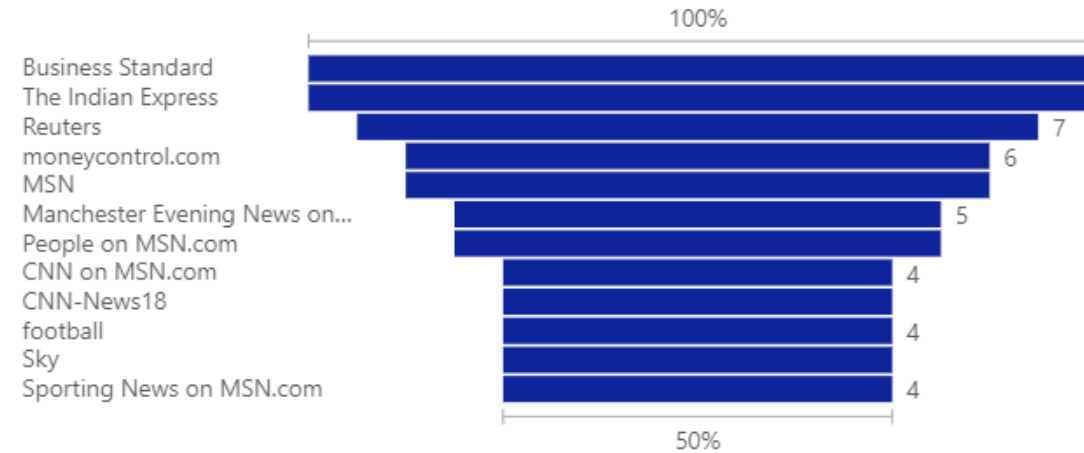


At 168, neutral had the highest Count of title and was 16,700.00% higher than mixed, which had the lowest Count of title at 1.

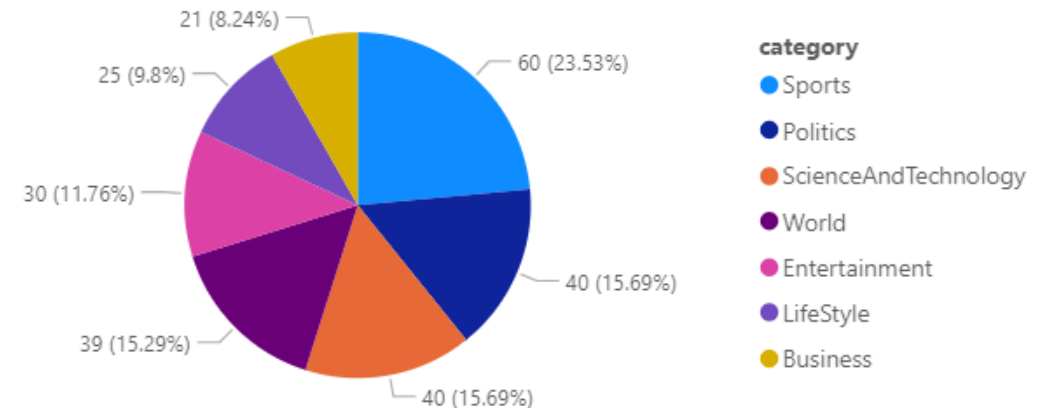
neutral had the highest Count of title at 168, followed by negative, positive, and mixed.

Across all 4 sentiment, Count of title ranged from 1 to 168.

Top 10 News Providers by count of news by provider



News items by category



Bing News Sentiment

10.87%
Positive

67.39%
Neutral







21.74%
Negative

46
Records

datePublished

10 Oct 2024



image	title	provider	url	category	datePublished
	Israel-Gaza-Lebanon live updates: IDF presses new north Gaza campaign	ABC	🔗	World	10 Oct 2024
	When will Davante Adams return? Latest injury updates on Raiders WR	AOL	🔗	Sports	10 Oct 2024
	Chinese trade restrictions on Australian lobsters to be lifted before Chinese New Year	Australian Broadcasting Corporation on MSN.com	🔗	Business	10 Oct 2024
	Cineworld pulls out of new leisure park	BBC	🔗	Business	10 Oct 2024
	Salisbury Poisonings - New BBC Podcast follows Public Inquiry into death of Dawn Sturgess	BBC	🔗	ScienceAndTechnology	10 Oct 2024
	Pictures show Paradise in Birmingham city centre as it gears up for flurry of new openings	Birmingham Mail	🔗	Entertainment	10 Oct 2024
scroll down to see more					

News Sentiment count by News Category

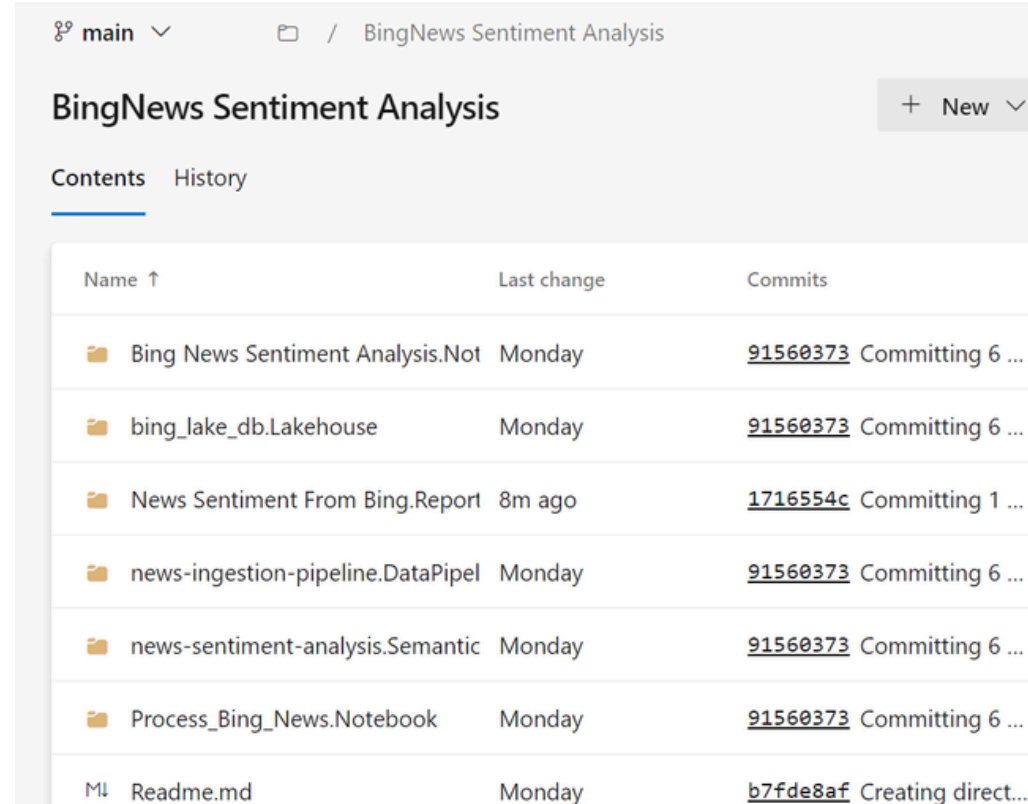
category	mixed	negative	neutral	positive	Total
Business		2	17	2	21
Entertainment	1	3	21	8	33
LifeStyle		5	17	3	25
Politics		6	32	4	42
ScienceAndTechnology		3	34	3	40
Sports		13	35	13	61
World		25	13	1	39
Total	1	57	169	34	261

News Sentiment count by News Provider

provider	mixed	negative	neutral	positive	Total
3dnatives.com			1		1
ABC		1			1
AFP on MSN.com			1		1
Africanews English on MSN.com			1		1
Al Jazeera on MSN.com		1			1
AL.com on MSN.com			1		
al.com on MSN.com			1		
AL.com on MSN.com					1
al.com on MSN.com					1
Total	1	57	169	34	261

10. Create Azure Devops Project repository and connect to Fabric Workspace

Sync the project - commit to the Azure repo



The screenshot shows the 'BingNews Sentiment Analysis' repository in Azure DevOps. The interface includes a breadcrumb path 'main / BingNews Sentiment Analysis', a '+ New' button, and tabs for 'Contents' and 'History'. A table lists the repository's contents with columns for Name, Last change, and Commits.

Name ↑	Last change	Commits
📁 Bing News Sentiment Analysis.Not	Monday	91560373 Committing 6 ...
📁 bing_lake_db.Lakehouse	Monday	91560373 Committing 6 ...
📁 News Sentiment From Bing.Report	8m ago	1716554c Committing 1 ...
📁 news-ingestion-pipeline.DataPipel	Monday	91560373 Committing 6 ...
📁 news-sentiment-analysis.Semantic	Monday	91560373 Committing 6 ...
📁 Process_Bing_News.Notebook	Monday	91560373 Committing 6 ...
📄 README.md	Monday	b7fde8af Creating direct...

12. Deploy to other environments

- a. Can create two more workspaces in same Fabric account for 'Test' and 'Prod' environments.
- b. Can create respective repositories in Azure DevOps and connect to fabric workspaces.
- c. Then can create deployment pipeline.

