



University
of Windsor

MILESTONE 2 Report

Jay Pandya: 110023841

Dhairya Chahuhan: 110025738

Shivang Thaker: 110013338

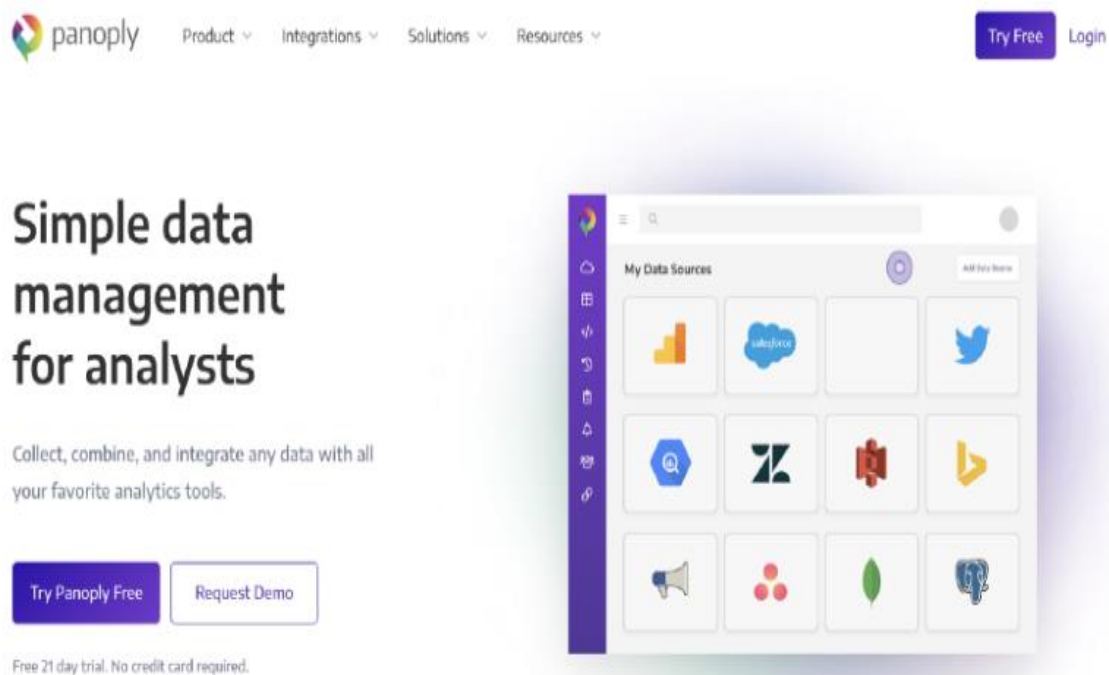
Agreta Gupta: 105182704

Submitted to: Prof. Kalyani Selvarajah

Instagram Insights using Business Intelligence

Steps which we have follow to continue after milestone 1:

Step 1: We have connected to Instagram graph API and collected Instagram data using APIs. We have used the data from an Instagram business account. Note that in milestone II we used data for testing. Next step is to collect our actual instagram data and show insights through MODE visualization tool.

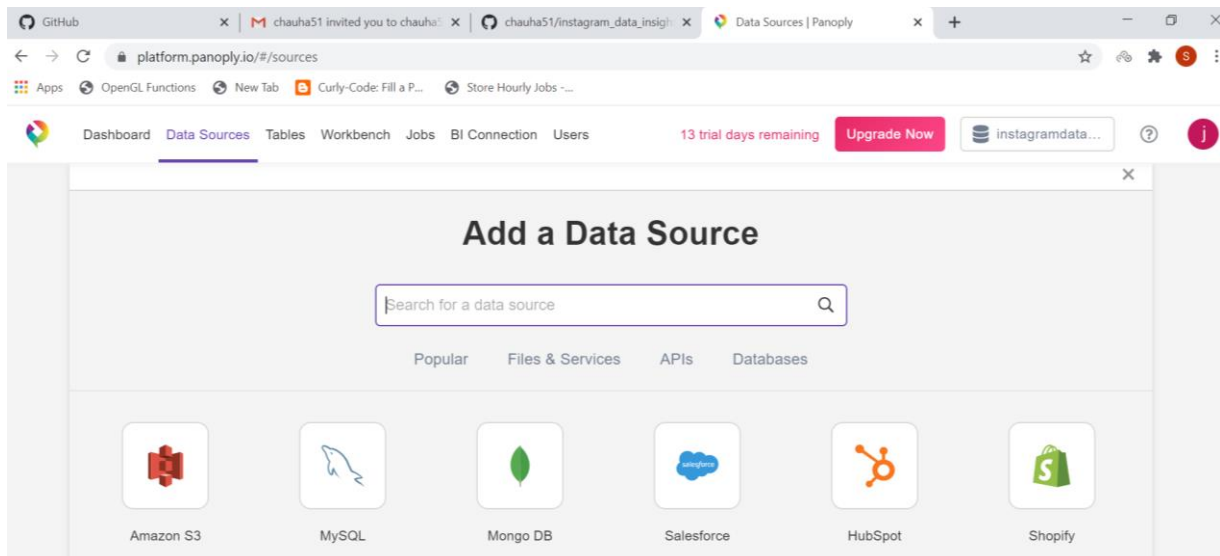


Advance Database Topics Group 8

Step 2: In this step, we collected data from different data sources to panoply warehouse, after putting credentials, we start to collect the data from the sources to panoply data warehouse.

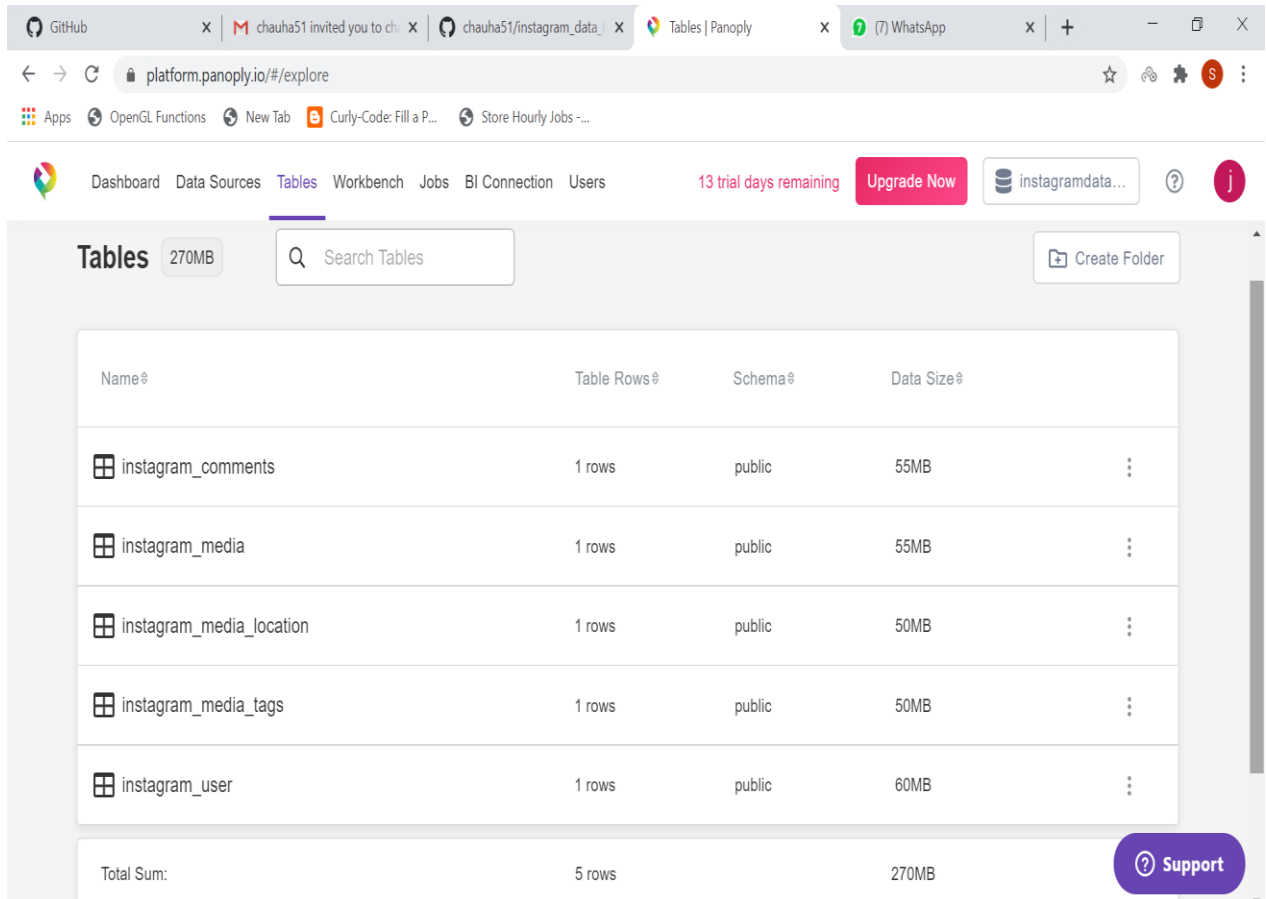
The screenshot shows the Panoply web interface for adding a new data source. The browser tabs include GitHub, an email, and the Panoply URL. The page title is 'Add a Data Source' and the breadcrumb is 'Data Sources > Amazon S3'. A notification says 'Changes made. Don't forget to save :)'. The 'General' tab is selected, showing fields for 'Address' (with an example: s3://mybucket/folder/file.tar.gz), 'Credentials' (AWS Access Key and AWS Access Secret), and a 'Next' button. Explanatory text on the right states: 'The prefix of the s3 files. We support many file formats, including csv, json and log files' and 'Your AWS credentials will be kept encrypted, but it's more secure to create a dedicated IAM user'.

Note: In our project we used instagram graph API as a data source.



Advance Database Topics Group 8

Step 3: Once we gathered data we can extract multiple tables from them, tables are shown below, note that here we make different tables of particular instagram user for testing. In final version we actually gives our team member business account dataset and throw query on it and analyse the results. Note : In actual data analysis there are multiple rows in one table.



The screenshot shows the Panoply web interface. The browser tabs include GitHub, an email invitation, a Panoply workspace, and WhatsApp. The address bar shows the URL `platform.panoply.io/#/explore`. The interface has a navigation bar with links to Dashboard, Data Sources, Tables (active), Workbench, Jobs, BI Connection, and Users. A trial notice indicates '13 trial days remaining' and an 'Upgrade Now' button is present. The main section is titled 'Tables' and shows a total size of '270MB'. A search bar is available. Below is a table listing the extracted tables:

Name	Table Rows	Schema	Data Size
instagram_comments	1 rows	public	55MB
instagram_media	1 rows	public	55MB
instagram_media_location	1 rows	public	50MB
instagram_media_tags	1 rows	public	50MB
instagram_user	1 rows	public	60MB
Total Sum:	5 rows		270MB

A 'Support' button is visible in the bottom right corner of the interface.

Advance Database Topics Group 8

Step 4: Now we have table so now we throw SQL query and see result of tables.

Here we are going to put analysis of our table.

The screenshot shows the Panoply Workbench interface. The left sidebar displays the 'public schema' with 'Tables' and 'Views'. The main editor contains a SQL query:

```
1 SELECT
2   instagram_media.id,
3   likes_count as "likes",
4   likes / cast(instagram_user.counts_followed_by as float) as "like engagement rate",
5   comments_count as "comments",
6   comments_count / cast(instagram_user.counts_followed_by as float) as "comments engagement rate"
7 FROM
8   public.instagram_media,
9   public.instagram_user
```

Below the query, the results are displayed in a table:

#	id	likes	like engagement rate	comments	comments engagement rate
1	abc1	234	0.241985522233713	34	0.03516028955532

The interface also shows a 'Run' button and a 'Support' link.

The screenshot shows the Panoply Workbench interface. The left sidebar displays the 'public schema' with 'Tables' and 'Views'. The main editor contains a SQL query:

```
1 with t AS
2 (SELECT
3   value as hashtag,
4   likes_count as likes,
5   comments_count as comments
6 FROM public.instagram_media m
7 left JOIN public.instagram_media_tags mt
8 ON m.id = mt.instagram_media_id
9 )
10 select
```

Below the query, the results are displayed in a table:

#	hashtag	avg_likes	avg_comments
1	nature	234	34

The interface also shows a 'Run' button.

Advance Database Topics Group 8

Step 5: Jobs are number of trials we did it records everything in panoply, how many time we succeed and failed

The screenshot shows the Panoply web interface. The top navigation bar includes links for Dashboard, Data Sources, Tables, Workbench, Jobs, BI Connection, and Users. A notification indicates '13 trial days remaining' and an 'Upgrade Now' button is present. The 'Jobs' section displays a summary of job statuses: Pending (0), Running (0), Successful (last 24 hours) (7), and Failed (last 24 hours) (5). Below this, a table lists individual jobs with columns for Name, Created, and Status. The jobs listed are 'File Upload (Manual collection)' with creation times of Aug 06, 2020 @ 00:03 UTC, Aug 06, 2020 @ 00:02 UTC, and Aug 06, 2020 @ 00:00 UTC, all with a 'Success' status. A 'Support' button is visible in the bottom right corner of the Jobs section.

Name	Created	Status
File Upload (Manual collection)	Aug 06, 2020 @ 00:03 UTC	Success
File Upload (Manual collection)	Aug 06, 2020 @ 00:02 UTC	Success
File Upload (Manual collection)	Aug 06, 2020 @ 00:00 UTC	Success

Step 6: Now in this step, we need to do BI connection (Business intelligence) . In this project we use Mode for Visualization. Here is credential for BI connection.

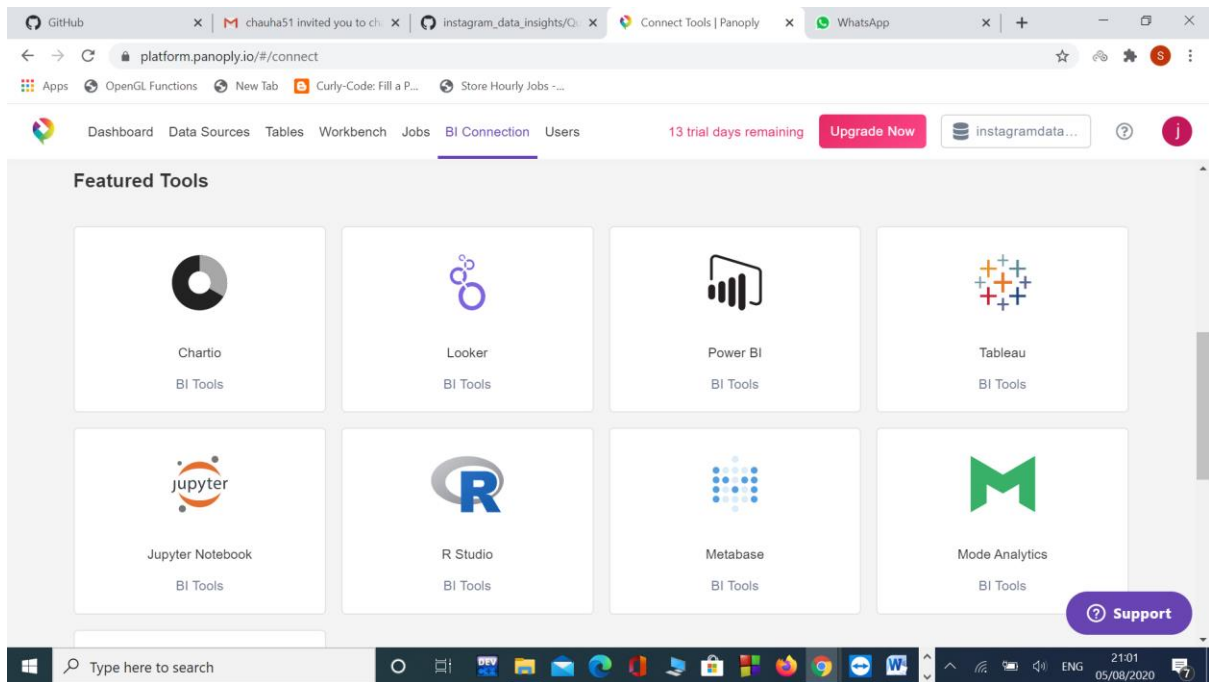
The screenshot shows the Panoply web interface with the 'BI Connection' tab selected. The 'Connect Tools' section displays 'Your BI Tool Connection Details' with the following information:

Field	Value
Host	db.panoply.io
Database	instagramdataanalysis
Port	5439
User	jaypandya083@gmail.com
Password	(same as your Panoply.io password)
Driver	Amazon Redshift or Postgres

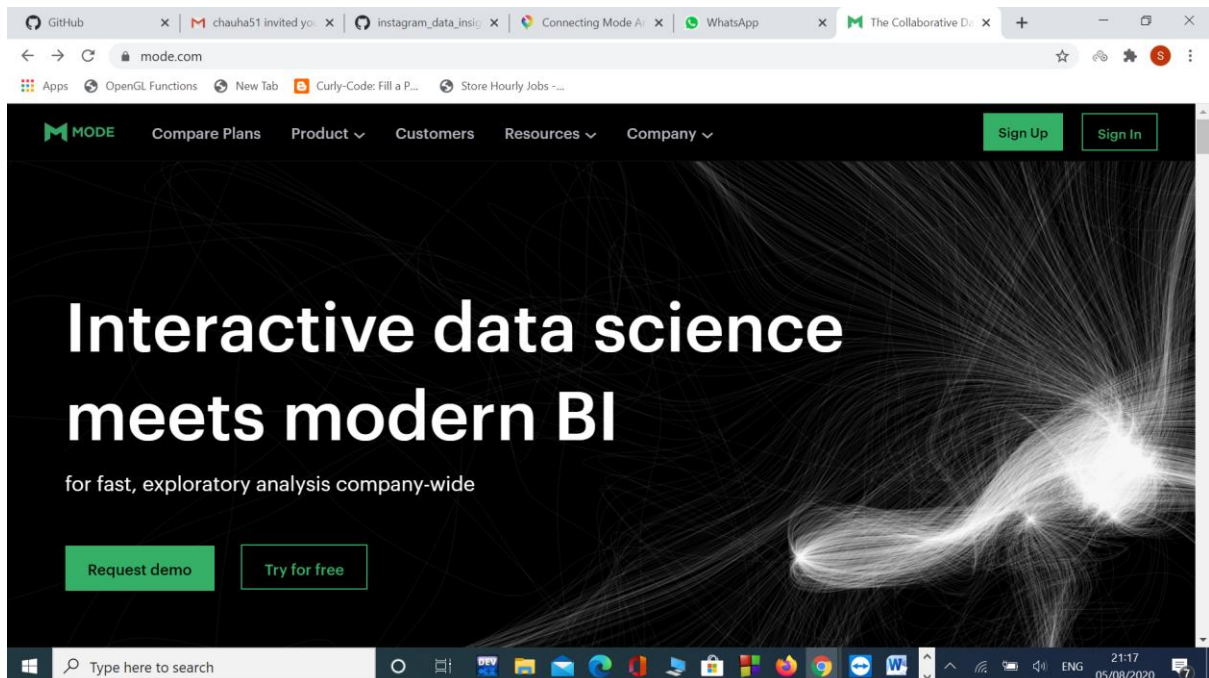
A 'Support' button is located in the bottom right corner of the Connect Tools section.

Advance Database Topics Group 8

These are BI tools, in which we use Mode



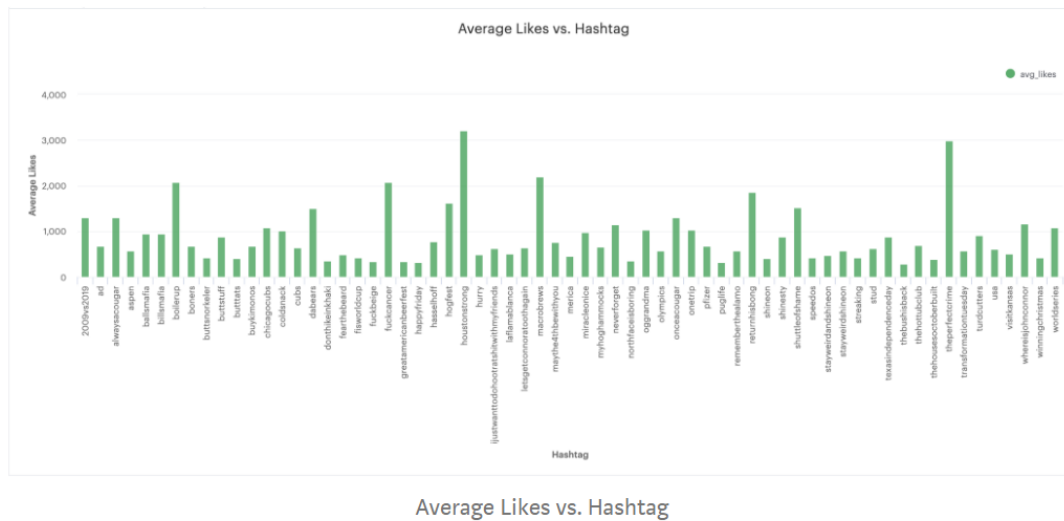
Step 7: After getting all data in Panoply database, our next step will connect Panoply to mode and fetch all the tables to account database in Mode. Once it is done, We can throw SQL query to analyse our graphical data in Mode.



Advance Database Topics

Group 8

In this mode we create account, and connect with panoply and we are able to see data in graphical form, we can see below,



We can follow same steps for each query and can see data in graphical form.

So, this is step by step explanation of our project.