

IBM PROJECT
SPRINT 1
(Team ID PNT2022TMID15458)

CONTENTS

1. Phase-1

- Downloading dataset from kaggle
- Extracting the dataset

2. Phase-2

- Logging into Cognos and IB DB2 service creation
- Logging into Cloud and IB DB2
- Connecting DB2 with Cognos

PHASE-1

a) Downloading dataset from Kaggle

The screenshot shows the Kaggle website interface. On the left is a sidebar with navigation links: Home, Competitions, Datasets, Code, Discussions, Learn, More, Your Work, and a 'RECENTLY VIEWED' section containing 'Google Play Store A...' and 'View Active Events'. The main content area features a search bar at the top. Below it, the dataset 'Google Play Store Apps' by user 'LAVANYA' (updated 4 years ago) is displayed. It has 4225 votes and a 'Download (2 MB)' button. The dataset description is 'Web scraped data of 10k Play Store apps for analysing the Android market.' Below the title are tabs for 'Data', 'Code (860)', and 'Discussion (81)'. The 'About Dataset' section includes a 'Context' paragraph explaining the data source. On the right, a 'Usability' score of 7.06 is shown, along with 'License' (Unknown) and 'Expected update frequency' (Not specified). A large Google logo is also visible on the right side of the dataset header.

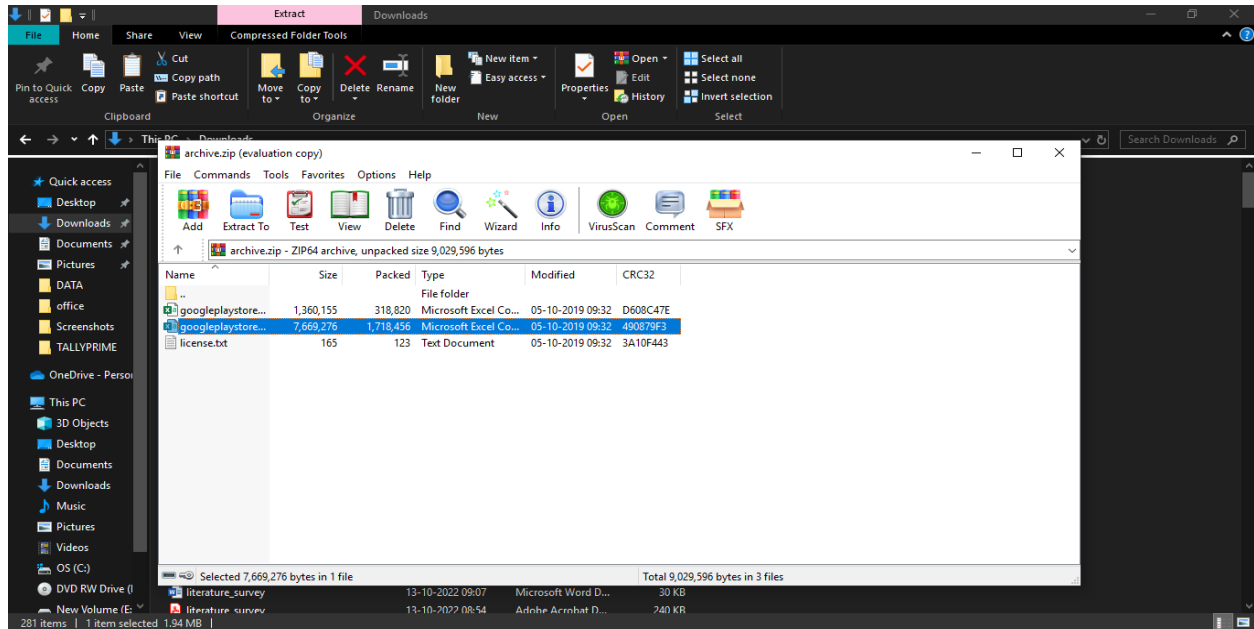
Google Play Store Apps
Web scraped data of 10k Play Store apps for analysing the Android market.

About Dataset

Context
While many public datasets (on Kaggle and the like) provide Apple App Store data, there are not many counterpart datasets available for Google Play Store apps anywhere on the web. On digging deeper, I found out that iTunes App Store page deploys a nicely indexed appendix-like structure to allow for simple and easy web

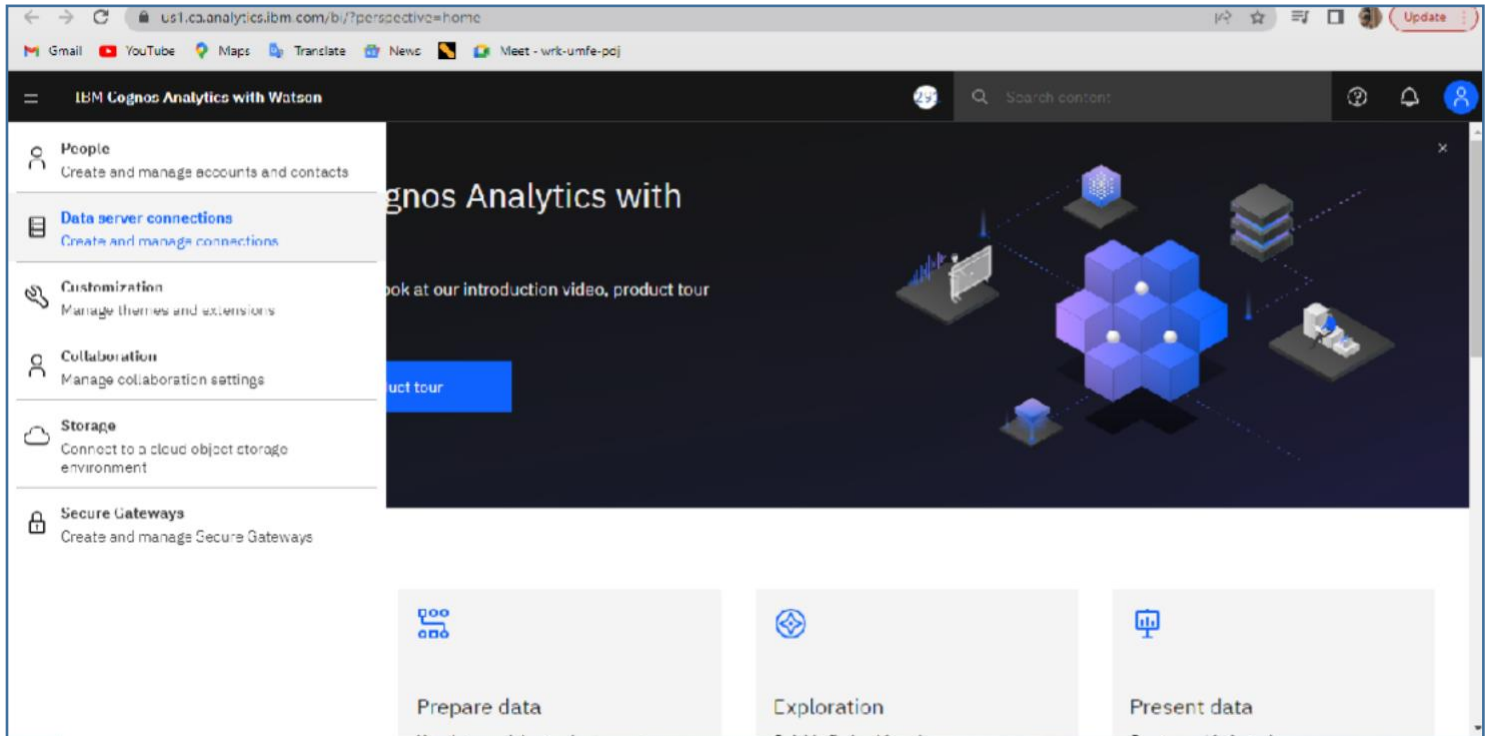
Usability 7.06
License Unknown
Expected update frequency Not specified

Extracting the dataset

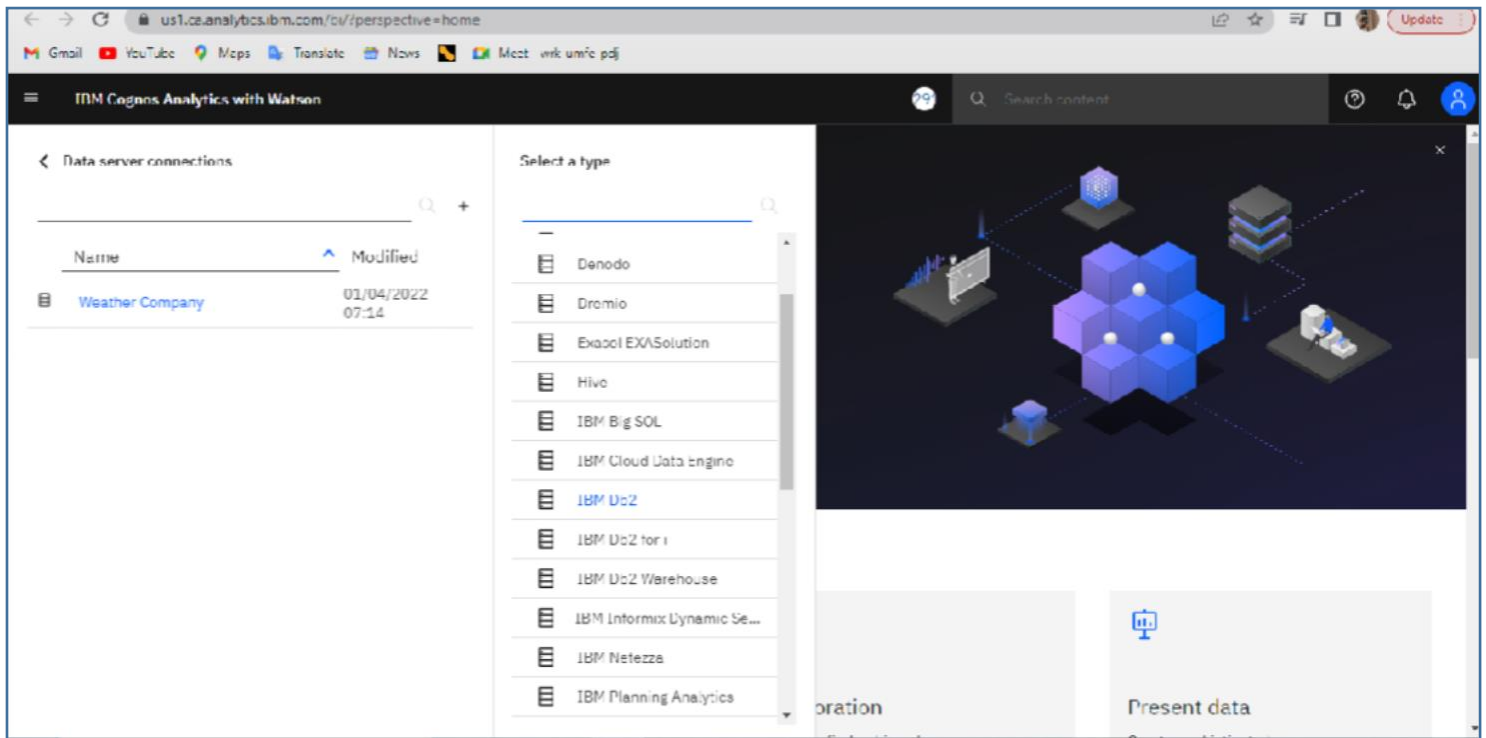


PHASE-2

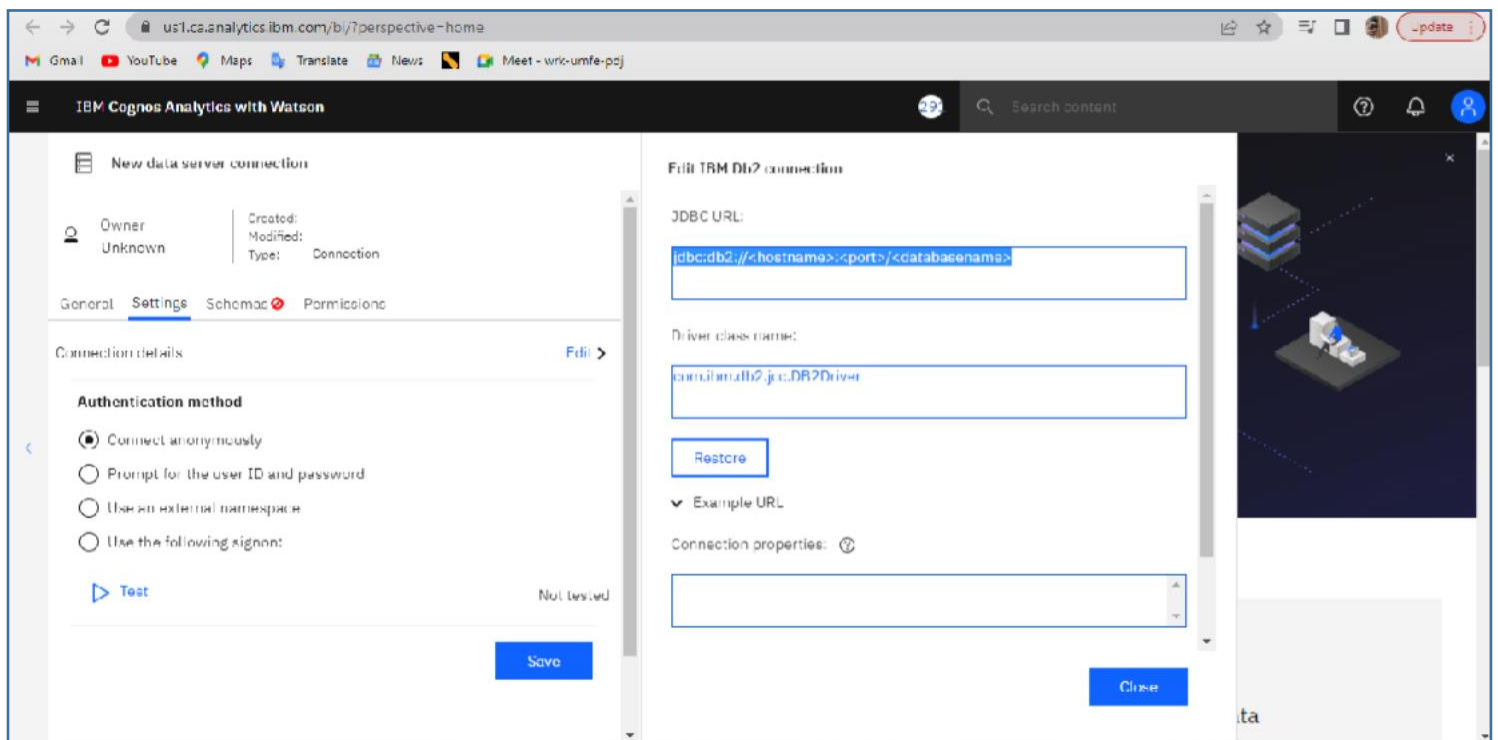
Establishing a IB DB2



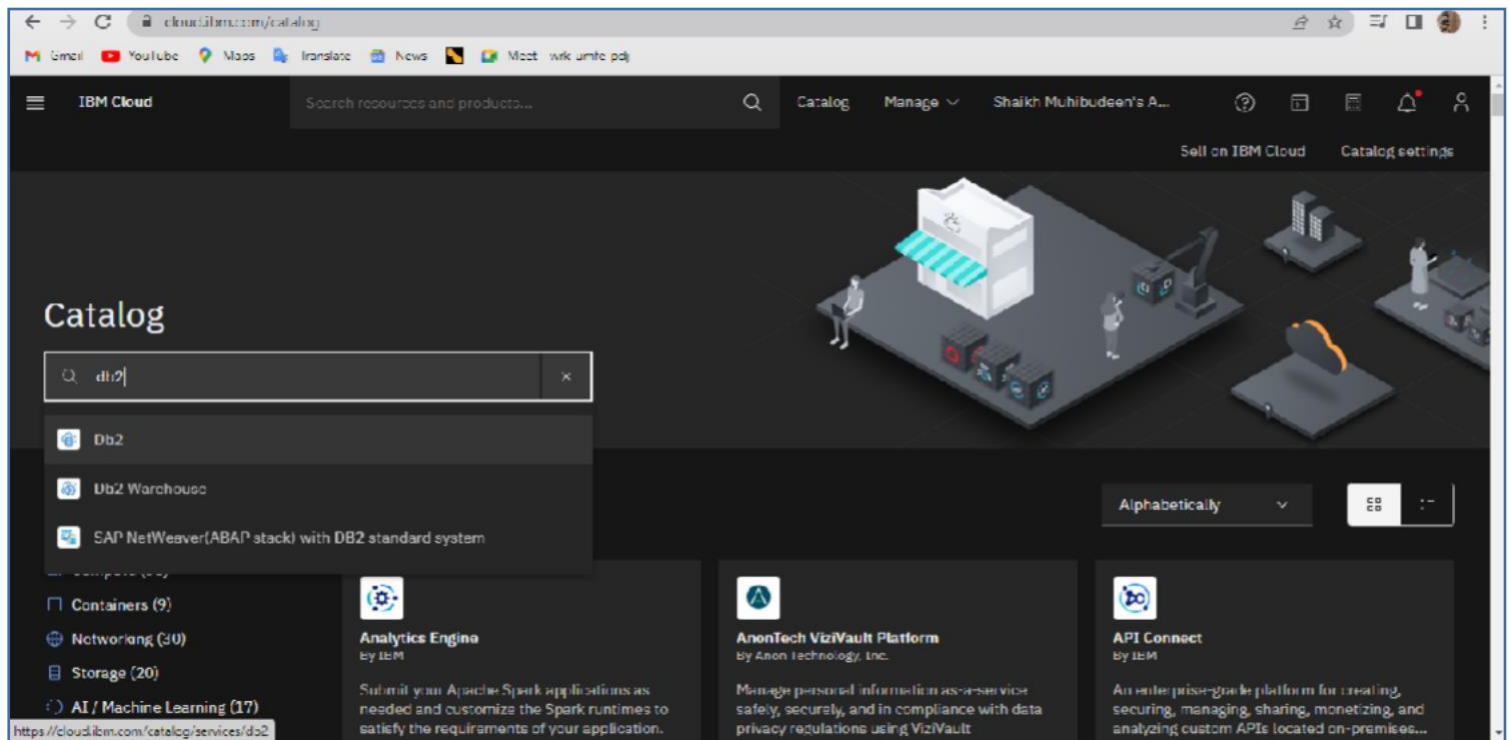
Selecting the type which is IBM Db2



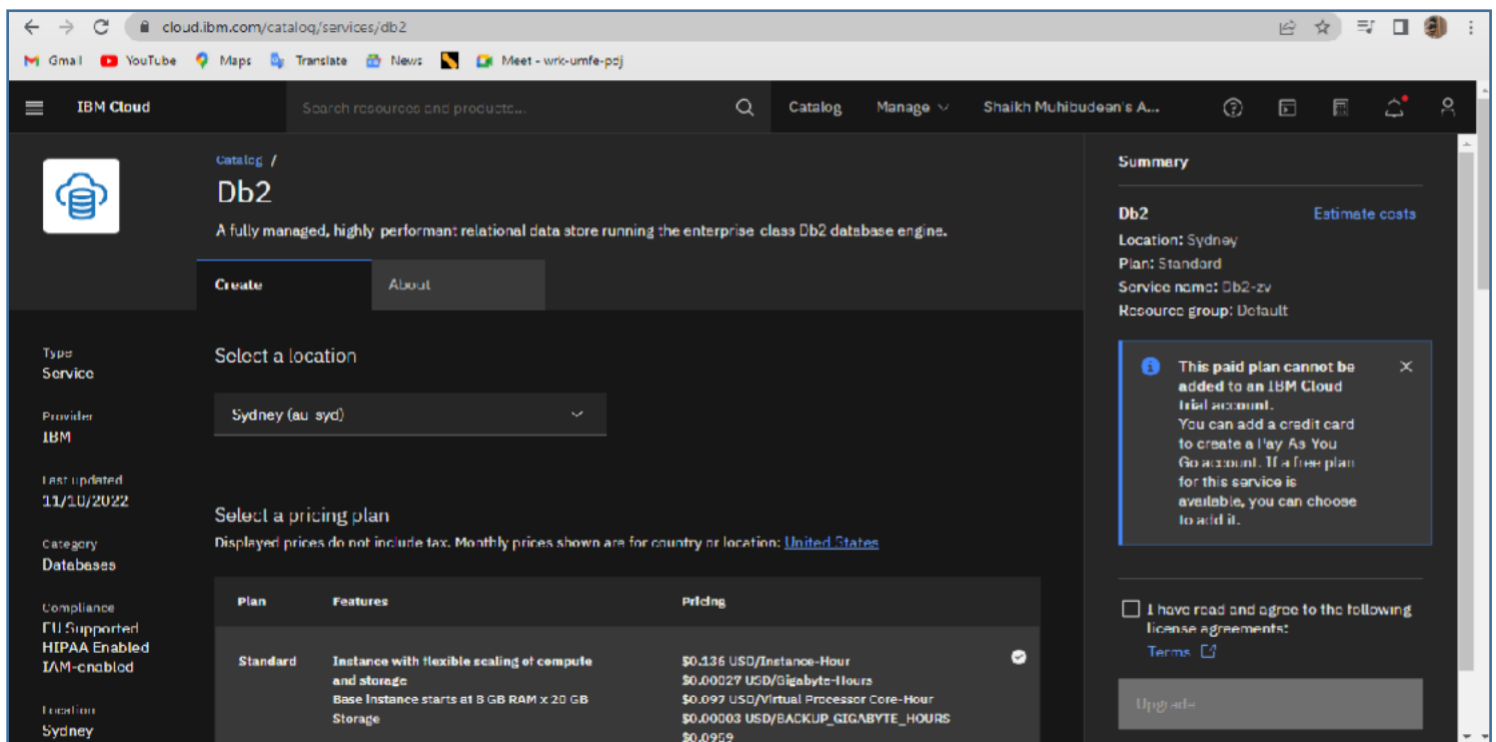
Changing Settings and Permissions



Logging into IBM cloud and searching Db2



Selecting region and Activating it



The Db2 is active

The screenshot shows the IBM Cloud console interface. At the top, the browser address bar displays the URL: `cloud.ibm.com/services/dashdb-for-transactions/cm%3Av1%3Abluemix%3Apublic%3Adashdb-for-transactions%3Aus-south%3Aa%2F7fe1526f52c24e3cb4e09697...`. The IBM Cloud logo is visible in the top left. A search bar and navigation links (Catalog, Manage) are in the top right. The main content area shows the resource **Db2-zv** with a green 'Active' status and an 'Add tags' link. A left sidebar contains a 'Manage' menu with options: 'Getting started', 'Service credentials', and 'Connections'. The main panel has two sections: 'Getting started' with instructions on finding credentials and a 'Go to UI' button, and 'Need help?' with a 'Support case' button.

Finding the Service credentials

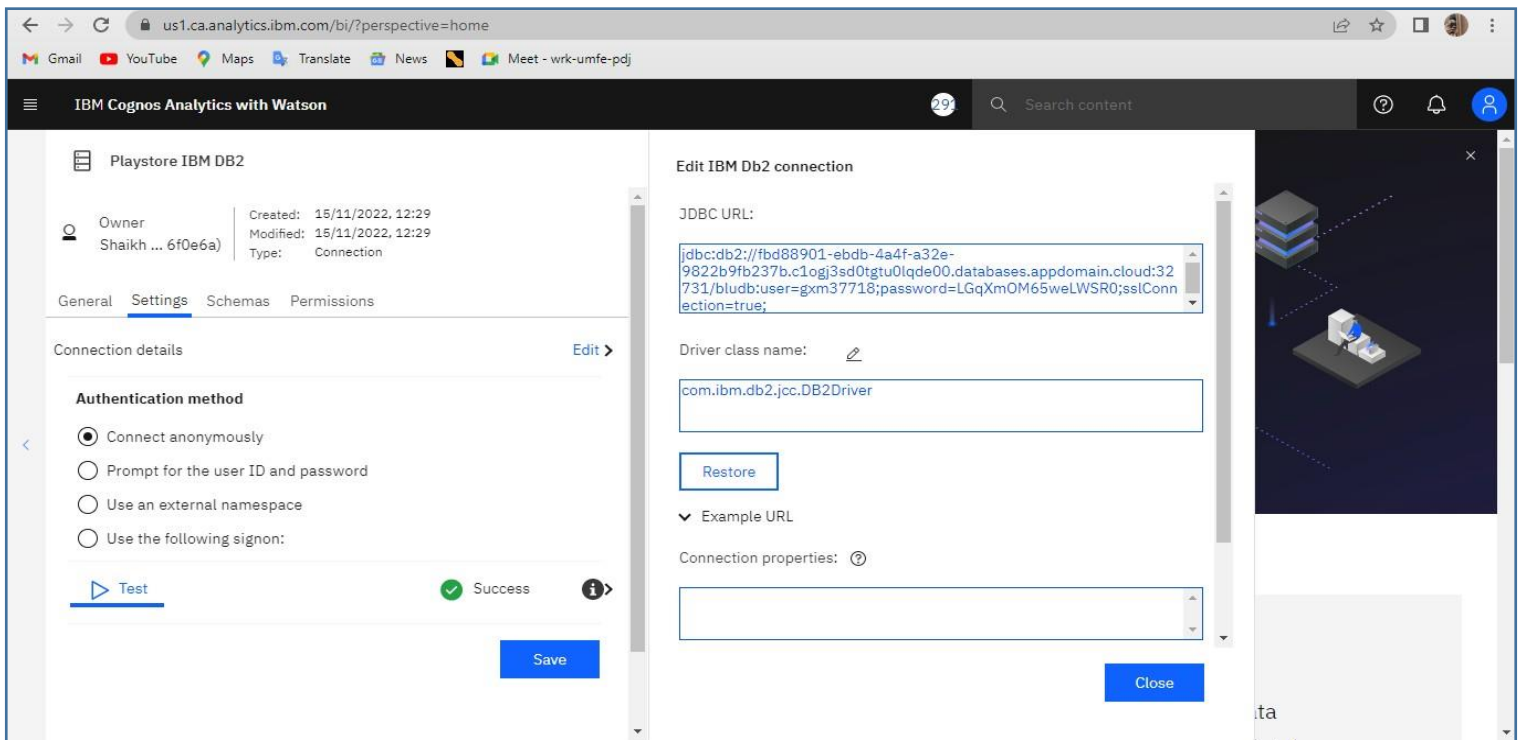
This screenshot shows the 'Service credentials' page for the **Db2-zv** resource. The left sidebar now highlights 'Service credentials'. The main panel features a search bar, a 'New credential' button, and a table of existing credentials. One credential, 'IBMCOGNOS', is listed with a creation date of '2022-11-15 10:06 AM'. Below the table, the JSON configuration for this credential is displayed.

Key name	Date created
IBMCOGNOS	2022-11-15 10:06 AM

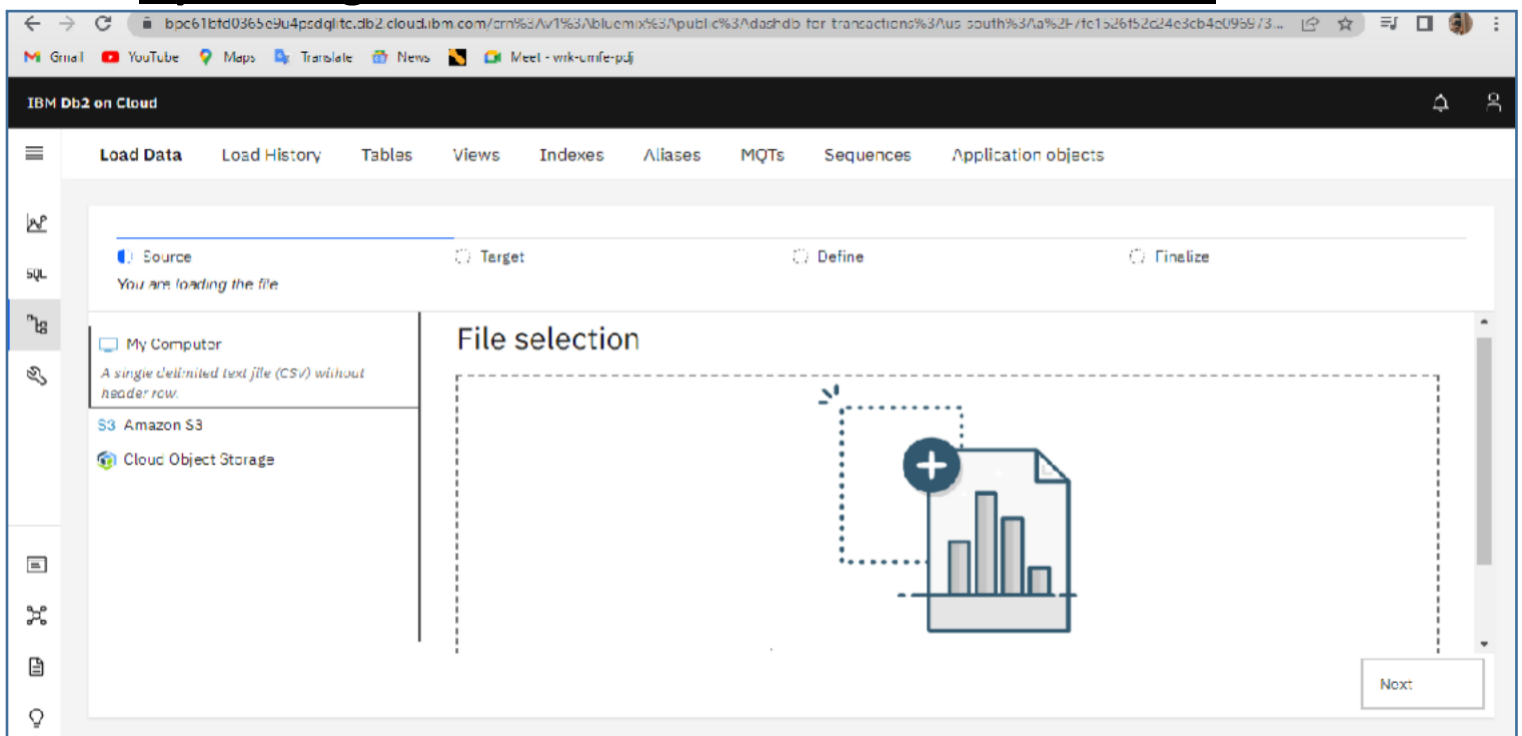
```
{
  "connection": {
    "cli": {
      "arguments": [
        "-u",
        "gxm07710",
        "-p",
        "LGqXm0MG5weLWSR0",
        "--ssl",
        "--sslCAFile",
        "1cbbb1b6-3a1a-4d49-9262-3102a0f7a7c0",
        "--authenticationDatabase",

```

Copying the URL from Cloud db2 as well as username and password to prevent difficulties while logging. Running Test



Uploading the downloaded file into the IBM Db2



Creating a Store Table

IBM Db2 on Cloud

Load Data | Load History | Tables | Views | Indexes | Aliases | MQTs | Sequences | Application objects

Source | Target | Define | Finalize

You are loading the file **googleplaystore (1).csv** into **GXM37718.STORE**

Schema

Find schemas

GXM37718 ✓

Table

Find tables in GXM37718

STORE ✓

Back | Next

Loading the csv file into Store data

IBM Db2 on Cloud

Load Data | Load History | Tables | Views | Indexes | Aliases | MQTs | Sequences | Application objects

Source | Target | Define | Finalize

You are loading the file **googleplaystore (1).csv** into **GXM37718.STORE**

Code page (character encoding): 1208 (UTF-8) | Separator: | Header in first row: ☒ | Time & date format: | Detect data types: ☐

APP	CATEGORY	RATING	REVIEWS
VARCHAR(129)	VARCHAR(19)	VARCHAR(3)	INTEGER
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159
2 Coloring book pages	ART_AND_DESIGN	3.9	967
3 101 Lashes Lila - FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644
5 Pixel Draw - Number Art: Coloring Book	ART_AND_DESIGN	4.3	967
6 Paper flowers instructions	ART_AND_DESIGN	4.4	167
7 Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178
8 Teddies Designer	ART_AND_DESIGN	4.1	24015

Back | Next

Finalizing the Loading Data process

The screenshot shows the 'Review settings' page in the IBM Db2 on Cloud console. The page is titled 'Review settings' and has a sub-header 'Summary'. Below the sub-header, there are several input fields for configuration: 'Code page' (1208 (Default)), 'Separator' (, (Default)), 'Time format' (HH:MM:SS (Default)), 'Date format' (YYYY-MM-DD (Default)), and 'Timestamp format' (YYYY-MM-DD HH:MM:SS (Default)). To the right of these fields, there is an 'Option' section with a label 'Maximum number of warnings' and a value of '1000'. At the bottom right, there are two buttons: 'Back' and 'Begin Load'.

Source Target Define Finalize

You are loading the file **googleplaystore (1).csv** into **GXM37718.STORE**

Review settings

Summary

Code page:	1208 (Default)
Separator:	, (Default)
Time format:	HH:MM:SS (Default)
Date format:	YYYY-MM-DD (Default)
Timestamp format:	YYYY-MM-DD HH:MM:SS (Default)

Option

Maximum number of warnings

1000

Back Begin Load


Running tests whether any errors are found

The screenshot shows the 'Status' page in the IBM Db2 on Cloud console. The page is titled 'Status' and has a sub-header 'Settings'. Below the sub-header, there is a large blue donut chart. To the right of the chart, there is a message 'The data load job succeeded. You can now work with your data.' Below the chart, there are three statistics: 'Rows read' (10,841), 'Rows loaded' (10,841), and 'Rows rejected' (0). Below these statistics, there are two timestamps: 'Start time' (11/16/2022 12:31:07 PM) and 'End time' (11/16/2022 12:31:28 PM). To the right of the main content, there is a section titled 'Errors' and 'Warnings'. The 'Errors' section shows '0' errors, and the 'Warnings' section shows '137' warnings. Below the 'Warnings' section, there is a message 'No errors. But, there are 137 warnings.'

WARNING 1 / warnings

googleplaystore (1).csv GXM37718.STORE

Status Settings



10,841 10,841 0

Rows read Rows loaded Rows rejected

Start time
11/16/2022 12:31:07 PM

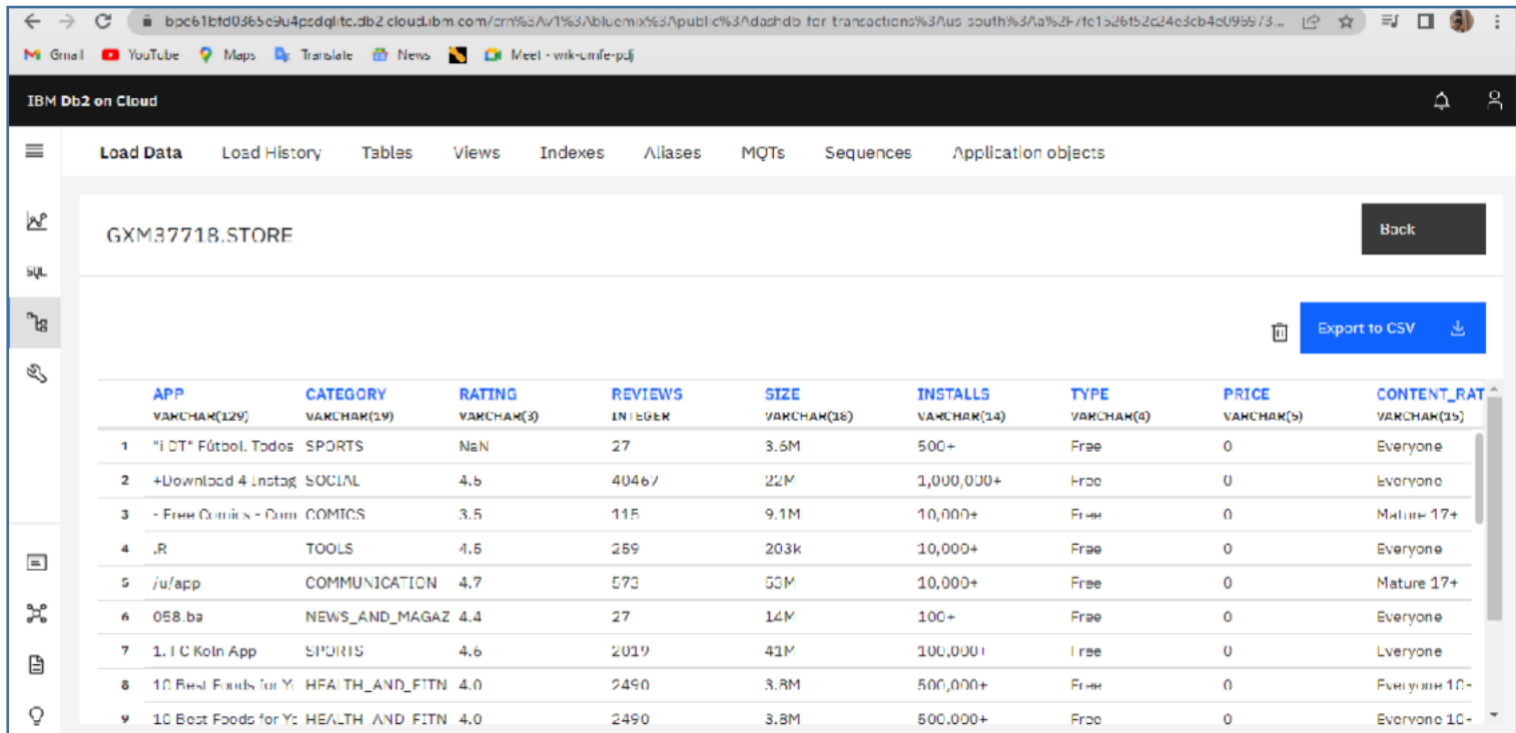
End time
11/16/2022 12:31:28 PM

The data load job succeeded.
You can now work with your data.

Errors 0 Warnings 137

No errors
But, there are 137 warnings.

Establishing a perfect Dataset using IBM Db2



IBM Db2 on Cloud

Load Data Load History Tables Views Indexes Aliases MQTs Sequences Application objects

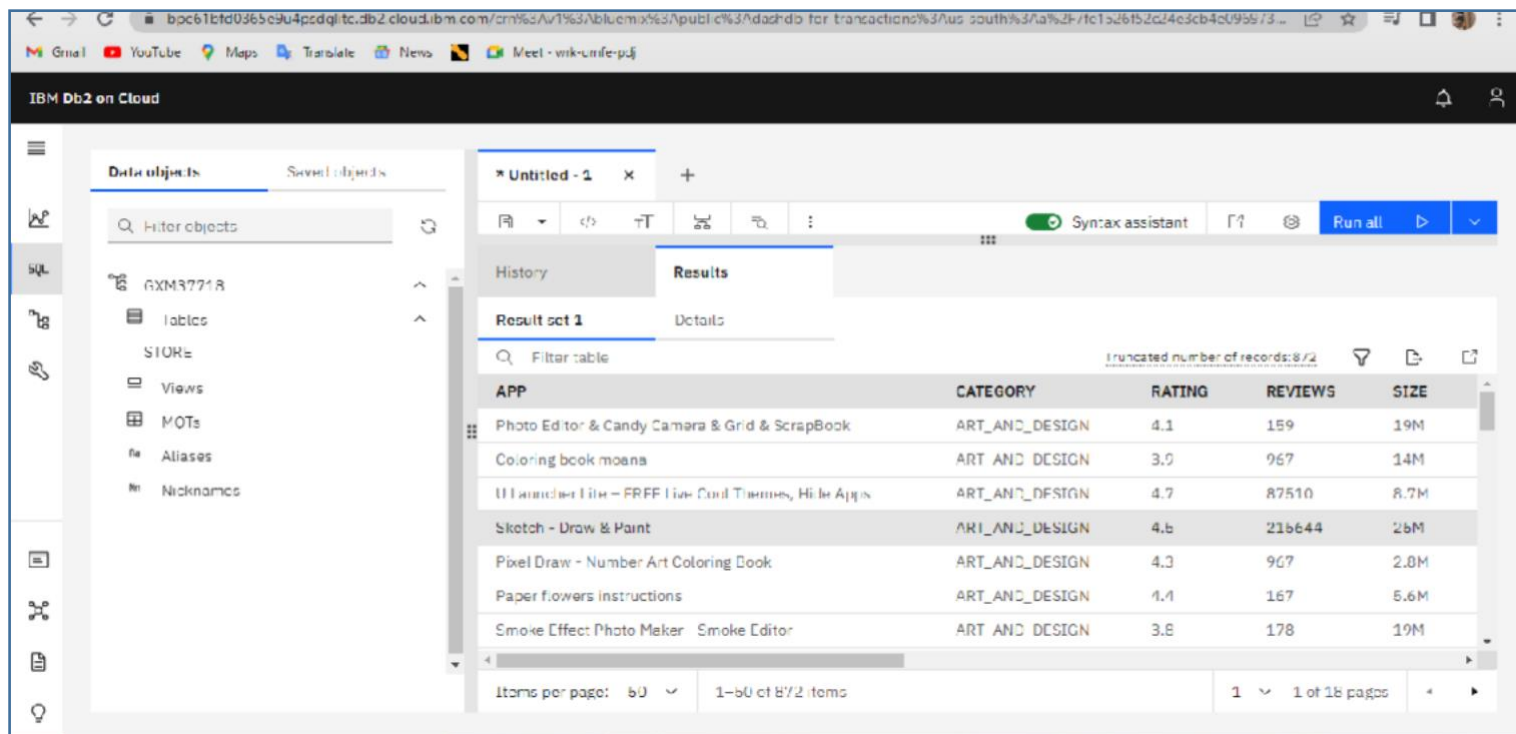
GXM37718.STORE

Back

Export to CSV

	APP VARCHAR(129)	CATEGORY VARCHAR(19)	RATING VARCHAR(3)	REVIEWS INTEGER	SIZE VARCHAR(18)	INSTALLS VARCHAR(14)	TYPE VARCHAR(9)	PRICE VARCHAR(9)	CONTENT_RAT VARCHAR(15)
1	"I DT" Fútbol Todos	SPORTS	NaN	27	3.6M	500+	Free	0	Everyone
2	+Download 4 Instag	SOCIAL	4.5	40467	22M	1,000,000+	Free	0	Everyone
3	- Free Comics - Com	COMICS	3.5	115	9.1M	10,000+	Free	0	Mature 17+
4	.R	TOOLS	4.5	259	203k	10,000+	Free	0	Everyone
5	/u/app	COMMUNICATION	4.7	573	53M	10,000+	Free	0	Mature 17+
6	058.ba	NEWS_AND_MAGAZ	4.4	27	14M	100+	Free	0	Everyone
7	1. I C Köln App	SPORTS	4.6	2019	41M	100,000+	Free	0	Everyone
8	10 Best Foods for Y	HEALTH_AND_FITN	4.0	2490	3.8M	500,000+	Free	0	Everyone 10+
9	10 Best Foods for Y	HEALTH_AND_FITN	4.0	2490	3.8M	500,000+	Free	0	Everyone 10+

Validating the Dataset roughly manually



IBM Db2 on Cloud

Data objects Saved objects

Filter objects

GXM37718

Tables

Views

MQTs

Aliases

Nicknames

*Untitled - 1

History Results

Result set 1 Details

Filter table

Truncated number of records: 8/2

APP	CATEGORY	RATING	REVIEWS	SIZE
Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M
Coloring book moana	ART_AND_DESIGN	3.9	967	14M
U Launcher Lite - FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M
Sketch - Draw & Paint	ART_AND_DESIGN	4.6	216644	26M
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.0M
Paper flowers instructions	ART_AND_DESIGN	4.1	167	5.6M
Smoke Effect Photo Maker Smoke Editor	ART_AND_DESIGN	3.8	178	19M

Items per page: 50 1-50 of 8/2 items 1 1 of 18 pages

Completing the DB2 connectivity with Cognos

The screenshot shows the IBM Cognos Analytics web interface. A blue banner at the top indicates "Maintenance: Scheduled maintenance completed. Click More Info for details and to subscribe to future events". The main content area is divided into two panels. The left panel, titled "Playstore IBM DB2", shows the connection details: Owner: Shaikh ... 6f096a, Created: 16/11/2022, 12:29, Modified: 16/11/2022, 12:29, Type: Data Server. The right panel, also titled "Playstore IBM DB2", shows the "Schemas" tab with a list of schemas: ERRORSHEMA, GXM37718 (selected), HEALINGMETRICS, TRMANT, and IBMCONSOLE. A table below the schemas shows the status and the number of tables loaded for each schema. The "GXM37718" schema is loaded with 1/1 tables. A "Load metadata" button is visible next to the selected schema. The background features a 3D visualization of data cubes.

Status	Schema name	Tables loaded
<input type="radio"/>	ERRORSCHEMA	
<input checked="" type="radio"/>	GXM37718	1 / 1
<input type="radio"/>	HEALINGMETRICS	
<input type="radio"/>	TRMANT	
<input type="radio"/>	IBMCONSOLE	

The Db2 is successfully connected to Cognos and is ready to use for exploration

The screenshot shows the IBM Cognos Analytics web interface. The top navigation bar includes "IBM Cognos Analytics with Watson" and "IBM Koggle". The main content area is divided into two panels. The left panel, titled "Data module", shows a search bar and a list of data modules: Navigation paths, Store, App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, and Last Updated. The right panel, titled "Grid", shows a "Preview data" button and a message: "To preview data, select a table, a column in a table, or a folder that contains columns." The background features a 3D visualization of data cubes.

Preview data

To preview data, select a table, a column in a table, or a folder that contains columns.