

Resources / Assignments (/COMP9321/20T1/resources/41975)

/ Week 7 (/COMP9321/20T1/resources/41978) / Assignment-2

Assignment-2

Specification

Make Submission

Check Submission

Collect Submission

Data Service for World Bank Economic Indicators

In this assignment, you are asked to develop a Flask-Restplus data service that allows a client to read and store some publicly available economic indicator data for countries around the world, and allow the consumers to access the data through a REST API.

IMPORTANT : *For this assignment , you will be asked to access the given Web content programmatically. Some Web hosts do not allow their web pages to be accessed programmatically, some hosts may block your IP if you access their pages too many times. During the implementation, download a few test pages and access the content locally - try not to perform too many programmatically.*

- The source URL: <http://api.worldbank.org/v2/> (<http://api.worldbank.org/v2/>)
- (<http://api.worldbank.org/v2/>) Documentations on API Call Structure:
(<https://datahelpdesk.worldbank.org/knowledgebase/articles/898581-api-basic-call-structure>)
<https://datahelpdesk.worldbank.org/knowledgebase/articles/898581-api-basic-call-structure>
(<https://datahelpdesk.worldbank.org/knowledgebase/articles/898581-api-basic-call-structure>)

(<https://datahelpdesk.worldbank.org/knowledgebase/articles/898581-api-basic-call-structure>)

The World Bank indicators API provides 50 year of data over more than 1500 indicators for countries around the world.

- List of indicators can be found at: <http://api.worldbank.org/v2/indicators>
(<http://api.worldbank.org/v2/indicators>)
- List of countries can be found at: (<http://api.worldbank.org/v2/countries>)
<http://api.worldbank.org/v2/countries> (<http://api.worldbank.org/v2/countries>)

As the documentation shows , you can construct URLs specifying different parameters to acquire necessary data. Use a custom URL to get information about a specific indicator. For example, the data on the GDP of all countries from 2012 to 2017 can be acquired via this URL:

(<http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017>)

[http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=json&per_page=1000)

[date=2012:2017&format=json&per_page=1000](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=json&per_page=1000)

([http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=json&per_page=1000)

[date=2012:2017&format=json&per_page=1000](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=json&per_page=1000))

For this assignment we are interested in annual values of a user specified indicator from 2012 to 2017 for one or all countries. The content of the file is quite straightforward. The data service specification will ask you to 'import' this data into a 'data storage'. You should inspect the content of the file carefully and decide on a data model and storage method. You will find that a JSON format is suitable to accessing data and publishing it.

Assignment Specification

The data service should use JSON format to publish its data and implement following operations.

Question-1: Import a collection from the data service (2.5 marks)

This operation can be considered as an on-demand 'import' operation. The service will download the JSON data for all countries (http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=json&per_page=1000) respective to the year 2012 to 2017 and identified by the indicator id given by the user and process the content into an internal data format and store it in the database; use **sqlite** for storing the data (the name of the database should be **YOURZID.db**).

For example, the following link can be used to obtain data for the indicator called: **NY.GDP.MKTP.CD**

(http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=json&per_page=1000)
[http://api.worldbank.org/v2/countries/all/indicators/ NY.GDP.MKTP.CD ?date=2012:2017&format=json&per_page=1000](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=json&per_page=1000)
(http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=json&per_page=1000)

To import data you require an indicator as a parameter given by the API user. The parameter should be a query parameter named:

- **indicator_id** : an indicator (<http://api.worldbank.org/v2/indicators>) <http://api.worldbank.org/v2/indicators> (<http://api.worldbank.org/v2/indicators>)

After importing the collection, the service should return a response containing at least the following information:

- **uri** : the URL with which the imported collection can be retrieved
- **id** : a unique integer identifier automatically generated
- **creation_time** : the time the collection stored in the database

Example:

```
HTTP operation: POST /collections?indicator_id=NY.GDP.MKTP.CD
```

Returns: 201 Created

```
{
  "uri" : "/collections/1",
  "id" : 1,
  "creation_time": "2019-04-08T12:06:11Z",
  "indicator_id" : "NY.GDP.MKTP.CD"
}
```

- You should return appropriate responses (JSON formatted) in case of already imported collections, invalid indicators or any invalid attempts to use the endpoint (e.g. If the input indicator id doesn't exist in the data source, return error 404)

Question 2- Deleting a collection with the data service (2.5 marks)

This operation deletes an existing collection from the database. The interface should look like as below:

```
HTTP operation: DELETE /collections/{id}
```

Returns: 200 OK

```
{
  "message" : "The collection 1 was removed from the database!",
  "id": 1
}
```

Question 3 - Retrieve the list of available collections (2.5 marks)

This operation retrieves all available collections. The interface should look like as like below:

```
HTTP operation: GET /collections?order_by={+id,+creation_time,+indicator,-id,-creation_time,-in
```



"order_by" is a comma separated string value to sort the collection based on the given criteria. Each segment of this value indicates how the collection should be sorted, and it consists of two parts (+ or -, and the name of column e.g., id). In each segment, + indicates ascending order, and - indicates descending order. Here are some samples:

+id,+creation_time	order by "id ascending" and then "creation_time ascending" In this case sorting by "id" has priority over "creation_time "
-creation_time	order by "creation_time descending"

Returns: 200 OK

```
[
  {
    "uri" : "/collections/1",
    "id" : 1,
    "creation_time": "2019-04-08T12:06:11Z",
    "indicator" : "NY.GDP.MKTP.CD"
  },
  {
    "uri" : "/collections/2",
    "id" : 2,
    "creation_time": "2019-05-08T12:16:11Z",
    "indicator" : "2.0.cov.Math.pl_3.all"
  },
  ...
]
```

Question 4 - Retrieve a collection (2.5 marks)

This operation retrieves a collection by its ID . The response of this operation will show the imported content from world bank API for all 6 years. That is, the data model that you have designed is visible here inside a JSON entry's content part.

The interface should look like as like below:

```
HTTP operation: GET /collections/{id}
```

Returns: 200 OK

```
{
  "id" : 1,
  "indicator": "NY.GDP.MKTP.CD",
  "indicator_value": "GDP (current US$)",
  "creation_time" : "2019-04-08T12:06:11Z"
  "entries" : [
    { "country": "Arab World", "date": 2016, "value": 2500164034395.78 },
    { "country": "Australia", "date": 2016, "value": 780016444034.00 },
    ...
  ]
}
```

Question 5 - Retrieve economic indicator value for given country and a year (2.5 marks)

The interface should look like as like below:

```
HTTP operation: GET /collections/{id}/{year}/{country}
```

Returns: 200 OK

```
{
  "id": 1,
  "indicator" : "NY.GDP.MKTP.CD",
  "country": "Arab World",
  "year": 2016,
  "value": 780016444034.00
}
```

Question 6 - Retrieve top/bottom economic indicator values for a given year (2.5 marks)

The interface should look like as like below:

```
HTTP operation: GET /collections/{id}/{year}?q=<query>
```

Returns: 200 OK

```
{
  "indicator": "NY.GDP.MKTP.CD",
  "indicator_value": "GDP (current US$)",
  "entries" : [
    {
      "country": "Arab World",
      "value": 2500164034395.78
    },
    ...
  ]
}
```

The <query> is an optional integer parameter which can be either of following:

- **+N (or simply N)** : Returns top N countries sorted by indicator value (highest first)
- **-N** : Returns bottom N countries sorted by indicator value

where N can be an integer value between 1 and 100. For example, a request like " GET /collections/1 /2015? query=+10 " should returns top 10 countries according to the indicator value.

Notice:


- Your code must implemented in **flask-restplus** and automatically generate swagger doc for testing the endpoints.
- Your code must be executable in CSE machines
- Your code must not require installing any software (python libraries are fine)
- Your code must be written in Python 3.5 or newer versions.
- Your operations (API methods) must return appropriate responses in JSON format, and appropriate HTTP response code! e.g., **500 Internal Server Error** is inappropriate response!
- Make sure you are using right datatypes in the database and in you API methods (e.g., not string for years '2017')

Submission:

- One and only one Python script file named " YOUR_ZID .py" which contains your code
- **How I can submit my assignment?**
Go to the assignment page click on the "Make Submission" tab; pick your files which must be named "YOUR_ZID.py". Make sure that the files are not empty, and submit the files together.
- **Can I submit my file after deadline?**
Yes, you can. But 25% of your assignment will be deducted as a late penalty per day. In other words, if you be late for more than 3 days, you will not be marked.

Resource created 26 days ago (Thursday 27 February 2020, 12:54:52 PM), last modified a day ago (Monday 23 March 2020, 12:23:48 PM).

Comments





 Add a comment



Rachel Hwang (/users/z5162384) about 5 hours ago (Tue Mar 24 2020 08:43:17 GMT+1100 (Australian Eastern Daylight Time))

Do the attribute in the response body have to be in order? Is the following acceptable?

Response body

```
{
  "entries": [
    {
      "country": "Caribbean small states",
      "value": 70684558866.8336
    }
  ],
  "indicator": "NY.GDP.MKTP.CD",
  "indicator_value": "GDP (current US$)"
}
```

I'm returning indicator, indicator_value and then entries but it shows in this order

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 4 hours ago (Tue Mar 24 2020 10:13:01 GMT+1100 (Australian Eastern Daylight Time))

Order of attributes is not important

Reply



Andrew Lei (/users/z5207966) about 13 hours ago (Tue Mar 24 2020 01:22:39 GMT+1100 (Australian Eastern Daylight Time))

Does the title string (next to get) need to include the query (such as what I wrote in green)? My api still uses the correct request url (as shown in the last box) and works as intended but I was wondering if this was required? (this is for Q3). I have the same question for the optional 'q=<query>' portion in Q6?

GET /collections ?order_by {string}

Parameters

Name	Description
order_by * required string (query)	+id,-indicator_id

Execute Clear

Responses

Response content type: application/json

Curl

```
curl -X GET "http://127.0.0.1:5000/collections?order_by=id&-indicator_id" -H "accept: application/json"
```

Request URL

```
http://127.0.0.1:5000/collections?order_by=id&-indicator_id
```

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 6 hours ago (Tue Mar 24 2020 08:31:56 GMT+1100 (Australian Eastern Daylight Time))

No; it should not be there. query parameters are shown in swagger in the Parameter section

Reply



Tianyu Zhang (/users/z5135099) about 15 hours ago (Mon Mar 23 2020 23:27:36 GMT+1100 (Australian Eastern Daylight Time)), last modified about 14 hours ago (Tue Mar 24 2020 00:31:17 GMT+1100 (Australian Eastern Daylight Time))

Hi,

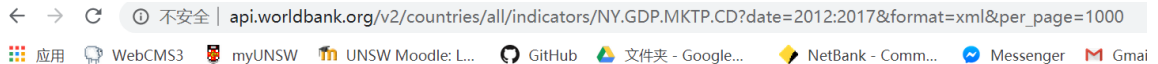
For questoin5, the example shows that the NY.GDP.MKTP.CD for Arab World in 2016 is 780016444034.00, but the value I found by GET

[http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=xml&per_page=1000)

[date=2012:2017&format=xml&per_page=1000](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=xml&per_page=1000)

([http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=xml&per_page=1000)

[date=2012:2017&format=xml&per_page=1000](http://api.worldbank.org/v2/countries/all/indicators/NY.GDP.MKTP.CD?date=2012:2017&format=xml&per_page=1000)) is 2510521024647.25. Is 780016444034.00 the standard answer for request in q5 example?



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<wb:data xmlns:wb="http://www.worldbank.org" page="1" pages="2" per_page="1000" total="1584" sourceid="2" lastupdated="2020-03-18">
  <wb:indicator id="NY.GDP.MKTP.CD">GDP (current US$)</wb:indicator>
  <wb:country id="1A">Arab World</wb:country>
  <wb:countryiso3code>ARB</wb:countryiso3code>
  <wb:date>2017</wb:date>
  <wb:value>2586506133266.57</wb:value>
  <wb:unit/>
  <wb:obs_status/>
  <wb:decimal>0</wb:decimal>
</wb:data>
  <wb:data>
    <wb:indicator id="NY.GDP.MKTP.CD">GDP (current US$)</wb:indicator>
    <wb:country id="1A">Arab World</wb:country>
    <wb:countryiso3code>ARB</wb:countryiso3code>
    <wb:date>2016</wb:date>
    <wb:value>2510521024647.25</wb:value>
    <wb:unit/>
    <wb:obs_status/>
    <wb:decimal>0</wb:decimal>
  </wb:data>
</wb:data>
```

And can we assume the data for requested indicator is already exist in the database?

And when I try to submit, it rejects my python file and shows this:

```
Session: 20T1

Submission: assign2
-----
No submission, all accept groups fail
-----

SUBMISSION ERROR:
Either you have not supplied all the required files or the files are not named correctly

Interpretation: assign2
-----
Accept any files matching *.hs
----- or -----
Accept 1 or more files matching *.c
and 0 or more other files matching *.h
and 0 to 1 other files matching [Mm]akefile
----- or -----
Accept 1 or more files matching *.java
and 0 or more other files matching *.htm
and 0 or more other files matching *.html
-----

Files supplied: assign2
-----
z5135099.py
-----

Submission NOT accepted
```

Is there any problem with submission program?

Thank you

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 6 hours ago (Tue Mar 24 2020 08:35:58 GMT+1100 (Australian Eastern Daylight Time))

Hi,

Examples are not based on real values.

No, if the data does not exist in database you must return "not found";
Regarding the submission, please wait for a day or so.

Reply



Corina Chen (/users/z5118831) about 15 hours ago (Mon Mar 23 2020 22:54:11 GMT+1100 (Australian Eastern Daylight Time)), last modified about 15 hours ago (Mon Mar 23 2020 22:54:20 GMT+1100 (Australian Eastern Daylight Time))

For Q6, when q is given are we sorting by the "value" field which is a float?

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 6 hours ago (Tue Mar 24 2020 08:36:59 GMT+1100 (Australian Eastern Daylight Time))

Yes

Reply



Xu Gao (/users/z5141401) about 16 hours ago (Mon Mar 23 2020 22:13:35 GMT+1100 (Australian Eastern Daylight Time))

For Q1, do we have to use `requests_cache` to do the insert or select SQL function for the database? I did not find a proper function to query records in the cached database so far. So I think I may be off track.

Are we allowed 3rd party modules like SQLAlchemy to set up our own database to store the results? Or can we save the results to a dataframe? Is it the expected way for this question?

Thanks for your time.

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 6 hours ago (Tue Mar 24 2020 08:39:03 GMT+1100 (Australian Eastern Daylight Time))

Why do not use the same approach as the first Lab activities to store the dataframe in the database? You can use pandas to store and load data.

I did not get why you want to cache?!

Reply



Andrew Lei (/users/z5207966) about 16 hours ago (Mon Mar 23 2020 22:09:45 GMT+1100 (Australian Eastern Daylight Time)), last modified about 16 hours ago (Mon Mar 23 2020 22:18:56 GMT+1100 (Australian Eastern Daylight Time))

Is there a default order (and number of outputs) for q6 when there is no query in the input? Also, do we just take 10 instead of +10 as a query input as it has to be an integer (+10 doesn't work unless it's a string). Is the creation_time given to us somewhere or is it something that I need to generate myself (datetime etc.) when I return the payload for Q1?

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 5 hours ago (Tue Mar 24 2020 08:40:47 GMT+1100 (Australian Eastern Daylight Time))

If q not provided, it should return all entries, and no need for ordering them. Creation time is the time that you fetched the indicator values and stored in the database

Reply



Shunyang Li (/users/z5139935) about 18 hours ago (Mon Mar 23 2020 20:04:12 GMT+1100 (Australian Eastern Daylight Time))

Hi for Q3, is the **order_by** argument fixed only can be **id**, **creation_time**, **indicator** or it can be any columns in the database?

If it can be any columns, does it have a fixed name? Because it's impossible for every one have the same columns name.

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 5 hours ago (Tue Mar 24 2020 08:41:54 GMT+1100 (Australian Eastern Daylight Time))

can be any valid column in the dataset.
And the dataset fields are already known.

Reply



Corina Chen (/users/z5118831) about 21 hours ago (Mon Mar 23 2020 17:08:00 GMT+1100 (Australian Eastern Daylight Time))

Suppose my current database have only some records of indicator_id=NY.GDP.MKTP.CD (because some have been deleted). Now if i perform

```
HTTP operation: POST /collections?indicator_id=NY.GDP.MKTP.CD
```

Will this POST be successful or return error message? Since it is an "already imported collection"

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 21 hours ago (Mon Mar 23 2020 17:21:40 GMT+1100 (Australian Eastern Daylight Time))

When you delete an indicator you are removing all of its entries.
This situation thus should not happen.

Reply



Xixia Su (/users/z5210542) a day ago (Mon Mar 23 2020 13:39:30 GMT+1100 (Australian Eastern Daylight Time))

I can not submit the python file because it require C or some other type files...

```
Interpretation: assign2 ----- Accept any files matchin
```

Reply



Rahil Agrawal (/users/z5165505) a day ago (Mon Mar 23 2020 13:57:54 GMT+1100 (Australian Eastern Daylight Time))

+1

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 5 hours ago (Tue Mar 24 2020 08:42:35 GMT+1100 (Australian Eastern Daylight Time))

Will fix it soon

Reply



Rachel Hwang (/users/z5162384) a day ago (Mon Mar 23 2020 13:30:47 GMT+1100 (Australian Eastern Daylight Time))

How do we test our code on CSE machines? I'm currently using Postman to test on my computer but I'm unsure how to ensure it runs properly on cse machines.

Reply



Lloyd Richardson (/users/z5188286) about 24 hours ago (Mon Mar 23 2020 14:25:29 GMT+1100 (Australian Eastern Daylight Time))

You can test using a browser on a CSE machine.

Open the following URL:

`http://127.0.0.1:5000/` (`http://127.0.0.1:5000/`)

Note: you may get a port conflict when running your service on CSE machine.

If this happens you can start your Flask application on a different port, e.g.

```
app.run(debug=True, port=5454)
```

Reply



Mohammadali Yaghoubzadehfard (/users/z5138589) about 5 hours ago (Tue Mar 24 2020 08:43:06 GMT+1100 (Australian Eastern Daylight Time))

Thanks for your complete answer

Reply

Load More Comments