Informatica Cloud API with Python and Google Datastudio

How to collecting Data Integration Log’s via API and exposing this with Google Datastudio

Roberto Amorim

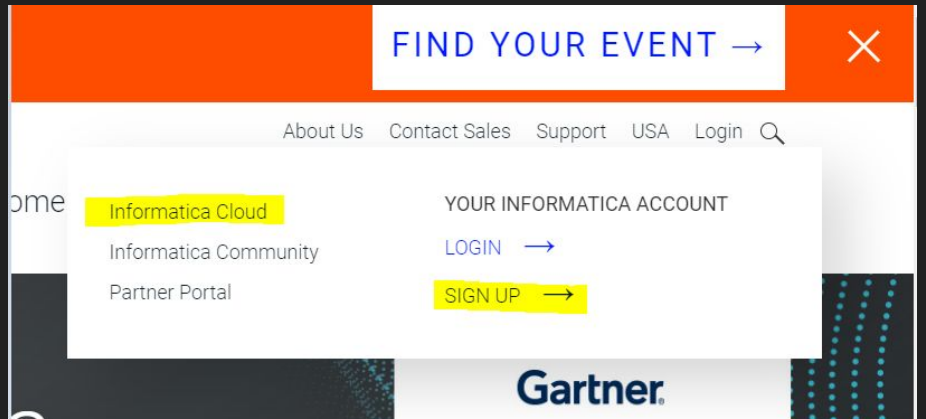
Informatica Rest API

Topics:

Rest API Reference is found in:

https://docs.informatica.com/integration-cloud/cloud-platform/current-version/rest-api-reference/informatica-intelligent-cloud-services-rest-api.html

It's necessary an authorized user to using this Rest API calls, but also it's possible to create a temporary account directly on site: informatica.com



Rest URLs

According to Rest API Reference, to use a 2o version API, we need to use a primary base URL to login and after logged in, with an Access Token, we gonna use the secondary base URL

- 1o base URL - `https://dm-us.informaticacloud.com/ma/api/v2/user/login`

If logon is successful, theresponse body has a field called ServerUrl , you must use this URL value in your calls now

in my example: `"serverUrl": "https://use4.dm-us.informaticacloud.com/saas"`

, but it's depends on which org regions you are

- so my 2o base URL - `https://use4.dm-us.informaticacloud.com/saas`

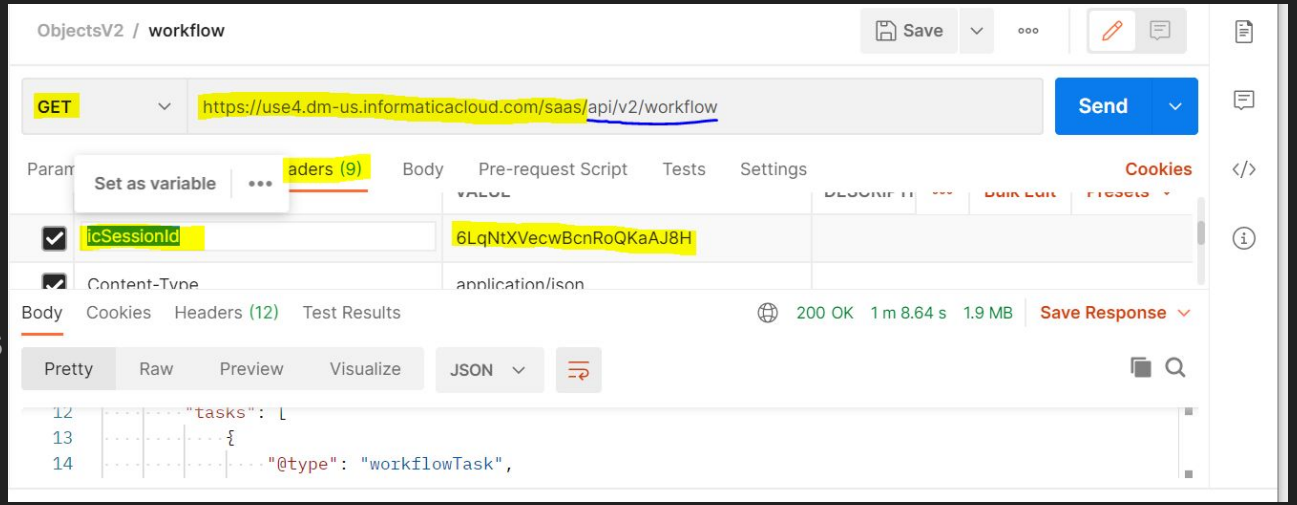
A good tool to test Rest API calls is **Postman** : https://www.postman.com/

Rest Methods

The first method is the login, is very useful to get these 2 principal fields:

serverUrl we talk about before and **isSessionId** which is a token to guarantee your valid connection to call Rest API Methods.

You must use this field **isSessionId** and your value in all REST API Calls (this is calling workflow method marked in blue) in Header with the same field name, as you can see in this example in **Postman**



REST API Methods used in this solution

We use some methods to get the entire view about the errors and warnings in execution in our organization, they are:

- /api/v2/workflow - list all metadata objects typed as a WorkFlow
- /api/v2/mttask/**007XYH0Z000000000AG5** - obligated to pass a object Id to search its details
- /api/v2/connection - Get information about all Connections Objects but not details as user name, onwer of schemas,
- /api/v2/connection/**007XYH0B0000000000XH** - obligated to pass the Connection Id to get its details
- /api/v2/activity/activityLog - to get Operations Logs of everything that is running your organization

Python

A Python Program is PreCalled from Informatica Cloud Integration Mapping Configuration, just to call Rest API, salve all information in a CSV file, than this Informatica Mapping is able to read that information and transfer to BigQuery Table

So we use

methods like

[example logon]:

```
def get_session_id(username, password):
    print("[get_session_id] inicio: " + " usuario: " + username + "; password: " + password)
    session_id = ''
    data = {'@type': 'connin', 'username': username, 'password': password}
    url = "https://dm-us2.informaticacloud.com/ma/api/v2/user/login"
    headers = {'Content-Type': 'application/json'}
    r = requests.post(url, data=json.dumps(data), headers=headers)
    print('[get_session_id]Codigo status API: ' + str(r.status_code))
    if r.status_code == 200:
        session_id = r.json()["icSessionId"]
        server_url = r.json()["serverUrl"]
    else:
        sys.exit(1)
    return session_id, server_url
```

Python

Or[example to get detail information of Mapping Task]:

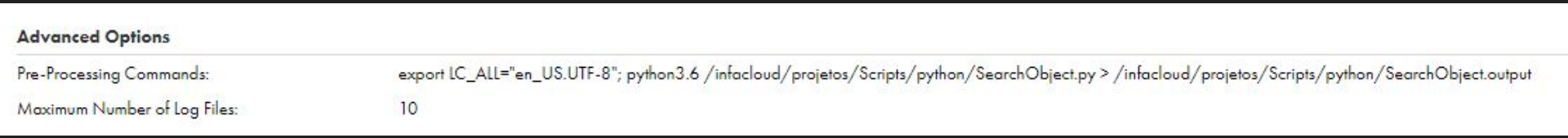
```
def SearchMetadataMCT(session_id, server_url, idMCT):
    print("[SearchMetadataMCT]inicio")
    job_start_url = server_url + "/api/v2/mttask/" + idMCT
    print(job_start_url)
    headers = {'Content-Type': 'application/json', 'icSessionId': session_id, 'Accept': 'application/json'
    data = {}
    r = requests.get(job_start_url, data=json.dumps(data), headers=headers)
    print("[SearchMetadataMCT]fim")
    return r.json()
```

Or [example to get details of Connection Object]:

```
def get_cnn_detail(session_id, server_url, connId):
    #print("[get_cnn_detail]inicio")
    job_start_url = server_url + "/api/v2/connection/" + connId
    headers = {'Content-Type': 'application/json', 'icSessionId': session_id, 'Accept': 'application/json'
    data = {}
    r = requests.get(job_start_url, data=json.dumps(data), headers=headers)
    #print("[get_cnn_detail]fim")
    return r.json()
```

IICS - ETL pre-execution

Informatica ETL is very useful, ease, intuitive and fast to create an integration, as you can see below, for every method Call, we use a python program, called from this mapping configuration



This is a bash command call to Python which extract and create a CSV File, you can use as output [> file.output]

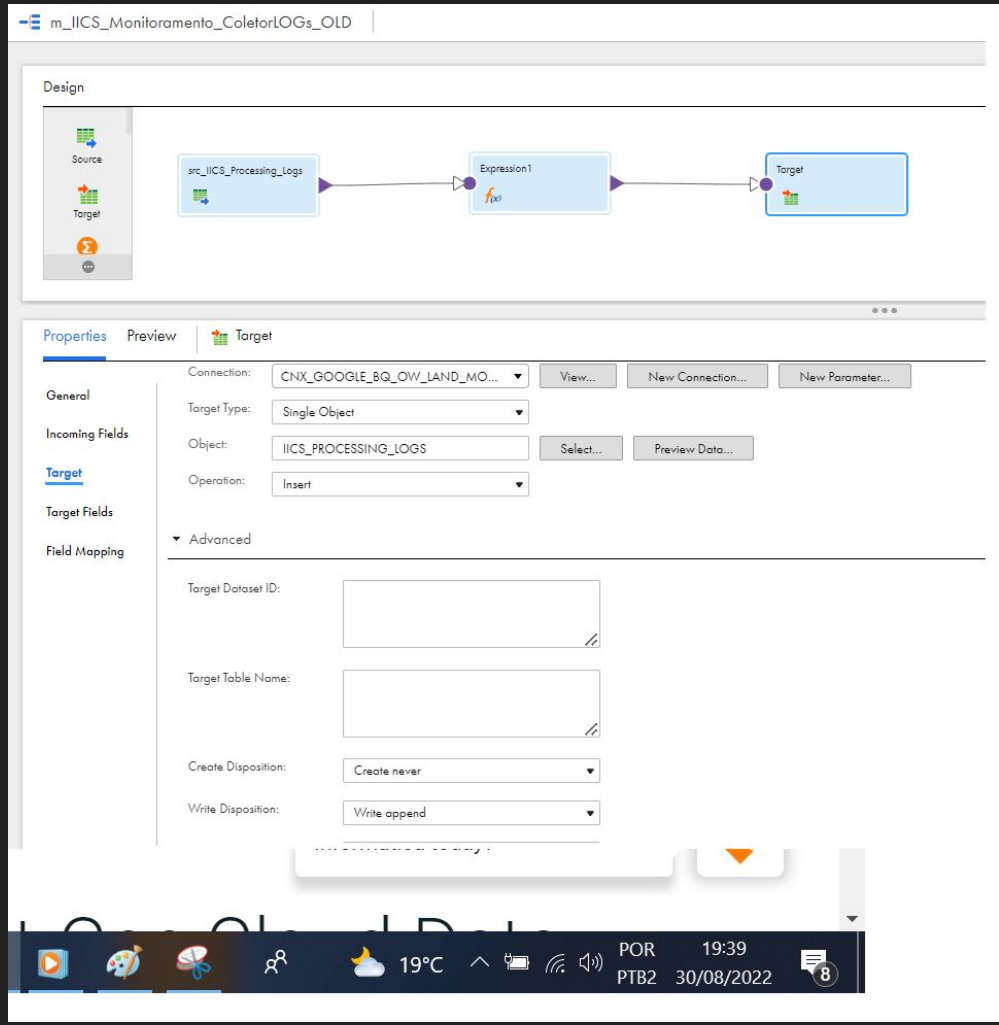
Informatica ETL BigQuery Load

In 3 steps Informatica do :

Read the CSV File [Python Result];

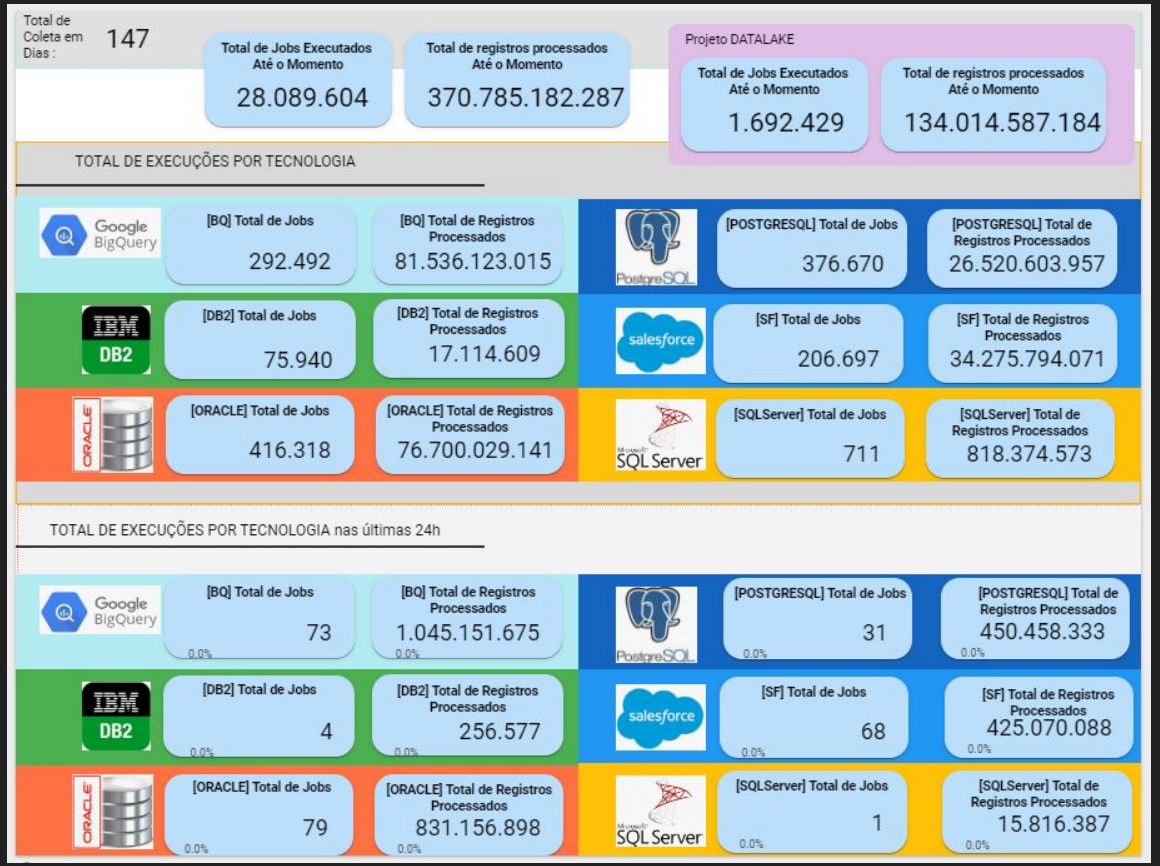
Rich field values, transform date into datetime or create new fields, as timestamp for example;

Feed in BigQuery Table, using Write Append mode, always Insert to create a historic of all running operations



Datastudio - using BigQuery Dataset as Datasource

As last, connecting Datastudio to the same Dataset of BigQuery table we can extracting Operation alerts, erros, executions and categorizing for technologies we can start to thinking about how to begin implement SRE - Site Reliability Engineering



linked-in: https://www.linkedin.com/in/roberto-amorim-580a1021b/

Engenheiro da Computação com Pós Gradução em Mecatrônica Industrial, sempre buscando automatizar operações. A frase de maior valor atualmente :

Making tomorrow better than today