CloudButton

CloudButton: Serverless Data Analytics

Pedro Garcia Lopez
Coordinator @ Universitat Rovira i Virgili

BIG DATA VALUE
PUBLIC-PRIVATE PARTNERSHIP

http://cloudbutton.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825184.

CloudButton

January 2019 - March 2022
4,277,507.50€

- The main goal is to create CloudButton: a Serverless Data Analytics Platform. CloudButton will "***democratize big data***" by overly simplifying the overall life cycle and programming model thanks to **serverless** technologies.
- To demonstrate the impact of the project, we target two settings with large data volumes: bioinformatics (**genomics**, **metabolomics**) and **geospatial** data (LiDAR, satellital).

EMBL

Atos

fundaciónMATRIX
10 años
INVESTIGACIÓN Y DESARROLLO SOSTENIBLE

The James Hutton Institute

IBM

UNIVERSITAT ROVIRA i VIRGILI

Imperial College London

Red Hat

Institut Mines-Télécom

Why CloudButton?

"Our proposal in this paper was motivated by a professor of computer graphics at UC Berkeley asking us "Why is there no **cloud button**?" He outlined how his students simply wish they could easily "push a button" and have their code – existing, optimized, single-machine code – running on the cloud."

E. Jonas. ***Occupy the Cloud: Distributed Computing for the 99%***

"What if every developer had access to the world's largest supercomputer?"

Tim Wagner. ***The Serverless SuperComputer****.*

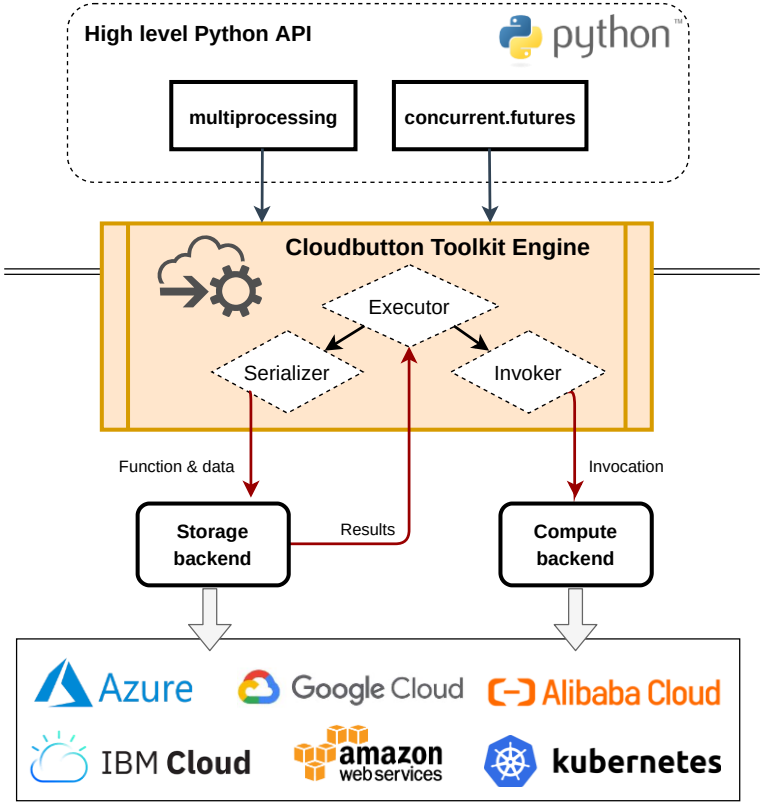
".. developers want the development experience of their laptops with the power of the cloud. They want an infinite laptop".

AnyScale.com. ***The Infinite Laptop****.*

" .. emerging programming models for an emerging computer architecture: the Cloud (Super) Computer. "

Rodric Rabbath. **The Dawn of the Cloud Computer**

LITHOPS



In this article, we advocate for **access transparency**: **enabling local and remote resources to be accessed using identical operations**. To this end, we present Lithops, a novel toolkit that allows regular, multiprocess Python programs to run at scale.

- Serverless Data Analytics in the IBM Cloud
- Serverless End Game: Dissaggregation Enabling Transparency
- Toward Multicloud Access Transparency in Serverless Computing

IBM

UNIVERSITAT ROVIRA i VIRGILI

Futures API

```python
from lithops import FunctionExecutor

def hello(name):
    return 'Hello {}!'.format(name)

with FunctionExecutor() as fexec:
    fut = fexec.call_async(hello, 'World')
    print(fut.result())
```

Multiprocessing API

```python
from lithops.multiprocessing import Pool

def double(i):
    return i * 2

with Pool() as pool:
    result = pool.map(double, [1, 2, 3, 4, 5])
    print(result)
```

Storage API

```python
from lithops import Storage

if __name__ == "__main__":
    storage = Storage()
    storage.put_object(bucket='my-bucket',
                       key='test.txt',
                       body='Hello World')
    ...
    print(storage.get_object(bucket='my-bucket',
                             key='test.txt'))
```

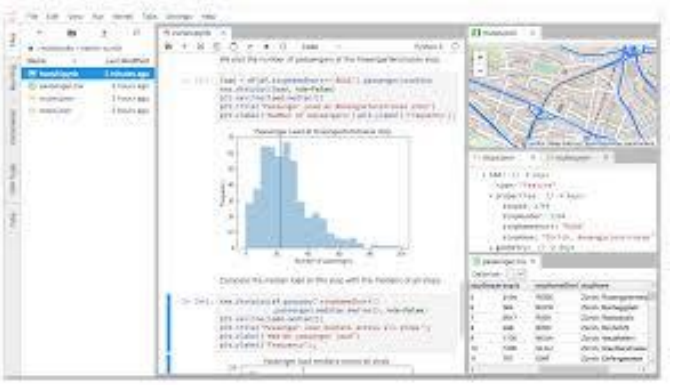
Storage OS API

```python
from lithops.storage.cloud_proxy import open,

if __name__ == "__main__":
    filepath = 'bar/foo.txt'
    with open(filepath, 'w') as f:
        f.write('Hello world!')

    dirname = os.path.dirname(filepath)
    print(os.listdir(dirname))
    os.remove(filepath)
```

Write once, run anyWhere

python

jupyterlab TNG



NumPy

pandas

scikit learn

OpenAI

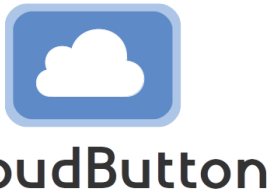
DASK

LITHOPS

IBM Cloud	aws	Microsoft Azure	Google Cloud	Alibaba Cloud	Red Hat	kubernetes
Cloud Functions Code Engine VPC Cloud Object Storage	AWS Lambda AWS S3	Functions Blob Storage	Cloud Functions Cloud Run Cloud Storage	Functions Compute Object Storage Service	OpenShift	Batch/Job Knative OpenWhisk OpenStack Swift Redis Ceph

CloudButton Hackathon

- Big Data Hackathon organized by URV in June 2021
- Using IBM technologies (IBM Cloud)
- Addressing Societal and Economic Impacts of COVID19 crisis
- Goals:
  - Generation of open text data sets (Web, social networks)
  - Preprocessing and generating structured data (pandas, csv)
  - Analytics and visualization using simple Jupyter Notebooks



THANK YOU!

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825184.