



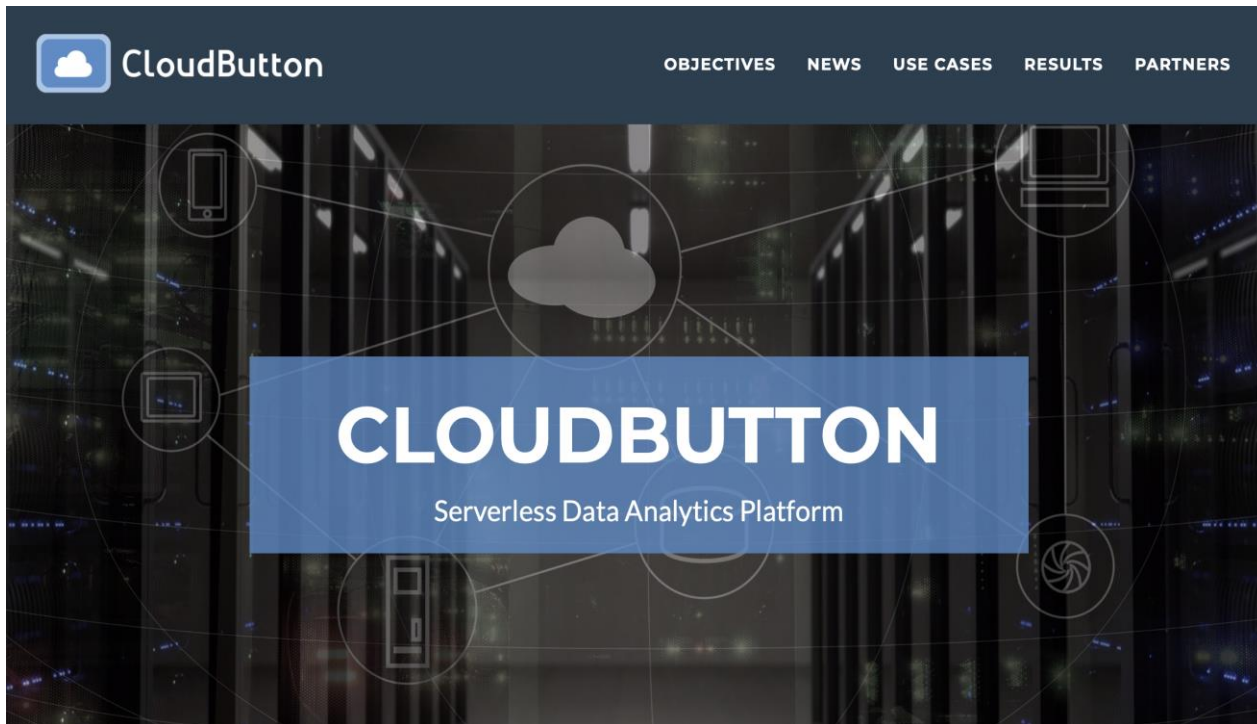
CloudButton

# Project Overview

Pedro Garcia Lopez

Coordinator @ Universitat Rovira i Virgili





<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 - June 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).



Background

# 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. “*

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

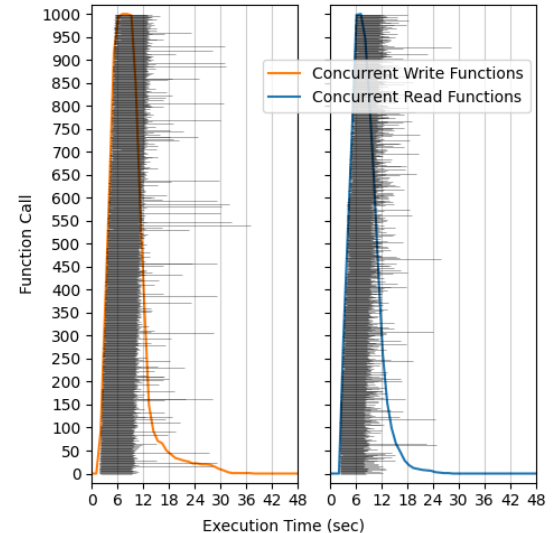
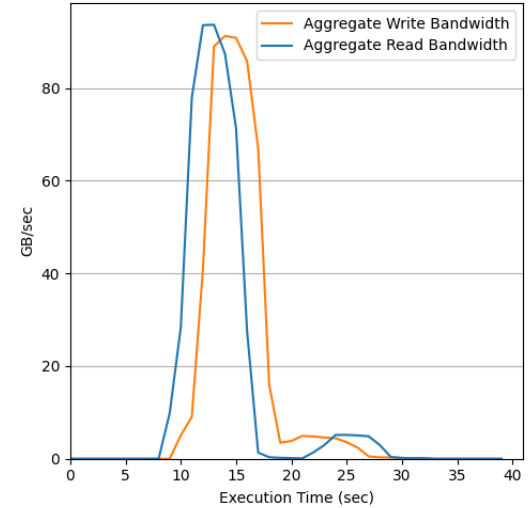
# Why Serverless ?

- Good for stateless embarrassingly simple parallel Jobs
- Auto-scaling and parallelism: 1000 parallel vCPUs and > 1000 GFLOPS
- > 100GB/s aggregate bandwidth between Serverless Functions and Cloud Object Storage
- Automatic provisioning, pay per use
- Elastic Resource Scaling



Amazon  
Lambda

- *Benchmarking parallelism in FaaS platforms. Future Gener. Comput. Syst. 124: 268-284 (2021)*
- *Lithops Multicloud FaaS Benchmark. 2002. <https://github.com/lithops-cloud/applications/tree/master/benchmarks>*



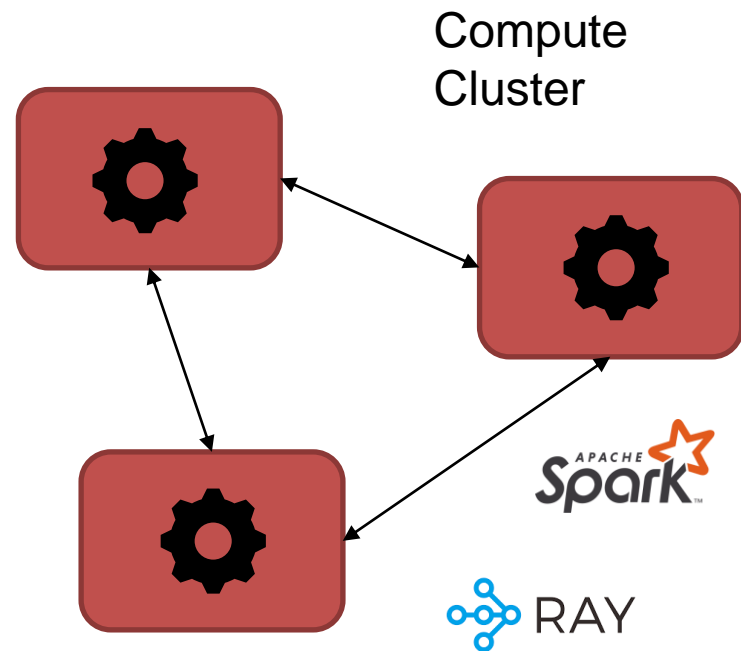
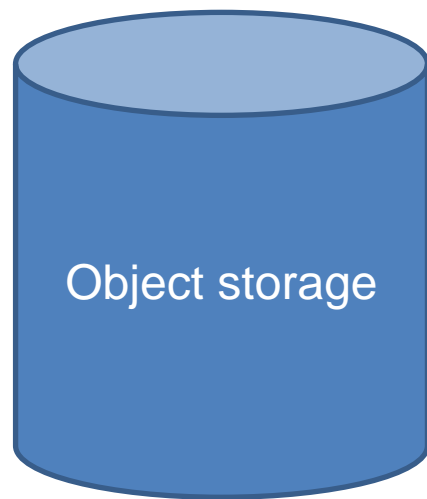
# Data Shipping model from Storage to Compute clusters

Ingest-then-Compute model

Massive Data Shipping

Parallel read/write to Storage is not always guaranteed with unstructured data

Compute clusters must be provisioned => overprovisioning or underprovisioning



# Serverless Model

**Data-driven model** -> Launch computing resources based on data volume

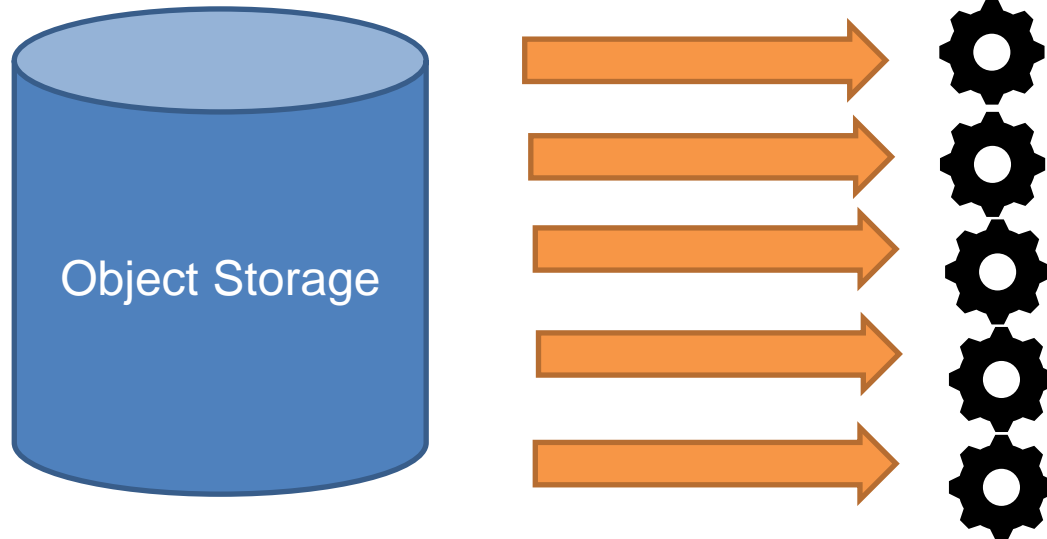
**Data Access partitioning** using objects or chunks (range)

Massive **Aggregate bandwidth**

**Parallel read** if unstructured data partitioning or cloud-optimized formats are available

**Parallel writes** with multipart uploads

**Elastic** Resource provisioning and **billing**



First review recommendations

# Recommendations

1. Please provide a plan in form of a memo for the 30/11/2020 proposing **urgent actions to mitigate the risk of not completing the validation** of the use cases by the end of the project. The experiments could be possibly refined. ✓ (Explained in Management)
2. Improve the testing and validation strategy of the use cases by **adding clear performance indicators**, in terms of **quantitative metrics or KPIs**, in order to precisely demonstrate the overall benefits and innovative aspects of the proposed solutions with respect to the current practice. Provide details in deliverable D2.5. ✓ (Explained in WP2, Project Overview)
3. Provide in deliverable D5.3 a **sound validation plan of the CloudButton Toolkit and involve also users external the consortium for its evaluation**. ✓ (Explained in WP2)
4. **Target more dissemination audiences** and ensure that all partners are actively participating in disseminating their results in the dissemination plan D6.3 ✓ (Explained in WP6)
5. Improve **the communication channels** of the project results, and report about it in the final review.
6. In deliverable D6.3 revise the **exploitation plan** of each partner, in particular the industrial ones, and include a business plan when possible. ✓ (Explained in Project Conclusions, WP6)
7. Report the final plan for **IPR management** in the deliverable D6.3. ✓ (Explained in WP6)

# Recommendations

1. Project's internal KPIs for the innovations that the project has produced in relation to the sota – beyond the knowledge that the partners previously had. ✓ (Explained in WP2)
2. What were the baselines in the case studies and what targets have been achieved.
  1. Metabolomics use case: baseline is previous Spark implementation vs Lithops ✓
  2. Genomics use case: baseline is HPC previous implementation and existing commercial solution (Illumina) ✓
  3. Geospatial use case: No previous implementation, baseline is speedup compared to sequential version, and comparison with different partitioned and not partitioned datasets. ✓
  4. Serverless Platform: compare to existing State of the Art technologies such as Spark, PyTorch, ScikitLearn or multiprocessing API in VMs and clusters of VMs. ✓
  5. User questionnaires to evaluate simplicity/productivity in WP2 led by ATOS ✓

# CloudButton KPIs/Goals

- **Simplicity/Productivity**
  - Front-end engineers and data analysts without Cloud knowledge
  - Hide Resource provisioning (data-driven approach)
  - Semi-transparent (Python notebooks) or fully transparent transition to the Cloud (CloudButton)
- **Performance**
  - Show performance improvements compared to cluster technologies like Spark
  - Low overheads in the transition to the Cloud
- **Scalability**
  - Proof that we scale to large data volumes using massive parallel computing power
  - Big Data pipelines (Use Cases)
- **Elasticity**
  - Demonstrate elastic workloads that benefit from the Serverless model
- **Cost**
  - Provide adequate cost/performance tradeoffs and offer alternatives services for batch analytics

# Project Outcomes and Impact

# Impact



CloudButton

## Four key Contributions (Innovation Radar)

- **Lithops** Serverless Data Analytics Platform
- EMBL's **Metaspace** Cloud Platform for Metabolomics
- RedHat's **Infinispan** In-Memory Data Grid
- **FaaSM** WebAssembly Runtime for Serverless Applications

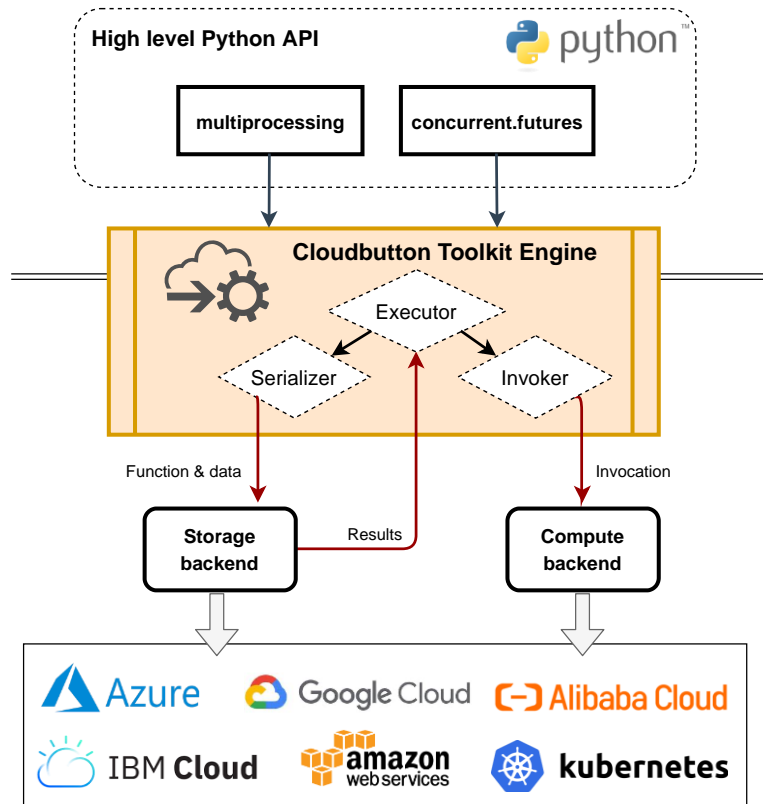
## Demonstrated KPIs in the different use cases

- **Simplicity/Productivity** in all use cases but METASPACE as flagship result
- Performance/Elasticity: parallelization of I/O and compute enabled huge speedups (46x)
- Better **performance** compared to commercial batch cloud solutions and HPC
- Validated in operational environment with hundreds of real users

## Dissemination

- High Scientific Impact (Top conferences and journals in Computer Science and BioInformatics)
- **Global Industrial Impact** (IBM Watson, Microsoft, Huawei, ATOS, SAP)
- Societal Impact (PyCon developer conference, Mass media publications, European Researchers night)
- More impact and results expected in the next months

# LITHOPS



**Lithops** is the major technical achievement of the CloudButton project. It is a **Multi-Cloud Serverless Data Analytic Platform** that is used in production at EMBL (Metaspace.eu), IBM, Hutton Institute, and Cyber ISCIII (DATOMA CLOUD)

Big Data pipelines in the **three use cases** (Genomics, Metabolomics, Geospatial) have **demonstrated all KPIs using Lithops** (IBM Cloud, AWS)

All partners have **integrated** their contributions with Lithops (Infinispan, FaasM)

Apache 2.0 license precludes any **IPR** future conflicts between partners

# Exploitation and next steps

Top conferences and journals. Industrial impact.  
Societal impact.

**EMBL's SpaceM startup** is incubation during 2022/23. Metaspace.eu is using Lithops in production in IBM Cloud with hundreds of users and large metabolomics datasets.

**Cyber ISCIll & URV's startup DATOMA Cloud** received different funds to build a startup in 2023. All Cloud developments are based on Lithops.

**IBM** is using Lithops in production with several clients, and it continues development and contributions to the repository. Josep Sampé (previously URV) now hired by IBM.

**Lithops** appears as infrastructure technology in at least 6 **Horizon Europe** projects (NearData, Extract, CloudSkin, CloudStars, Metacell, Metadrug). Four of them starting on January 2023.



**RedHat** is maintaining **Infinispan (RedHat Data Grid)**. A potential collaboration with IBM to deploy Metaspace/Lithops using OpenShift on premise locations like Astra Zeneca.

# Final Review

# Final Review Videos



CloudButton

## Overview

- CloudButton promotional video (RHT)
- Project overview and major outcomes (URV)

## Use cases

- Serverless Geospatial Data (URV)
- Serverless Genomics Pipelines (HTI)
- Serverless Metabolomics and METASPACE (EMBL, IBM)

## Transparency

- Efficient portage of Java and shell applications to serverless (IMT)
- Transparently running OpenMP/MPI applications on Serverless (IMP)
- Transparent Serverless execution of Python multiprocessing applications (URV)

## Integrations

- Supporting WebAssembly serverless functions with Lithops (IMP)
- Infinispan backend for Lithops (RHT)
- SLA Component and Lithops (ATOS)

# Final Review Agenda



**Meeting Subject:** CloudButton final review

**Venue:** Teams meeting

**Date:** September 15<sup>th</sup>, 2022

**Chair:** Pedro García López (Coordinator)

Time	Subject	Time (Mins)	Lead participant
9:00	Private meeting, reviewers and PO	30	PO
9:30	Welcome – Presentation of Participants	10	Coordinator
9:40	CloudButton promotional video	5	RHT
9:45	Overall Project Presentation Project Overview and Big Picture Achievements and project outcomes	15	Coordinator
10:00	WP2- Architecture and Use Cases	30	Coordinator, EMBL, Hutton
10:30	Round table and open questions*	20	PO
10:50	Break	10	
11:00	WP3- Serverless Compute Engine for Big Data	20	IBM
11:20	WP4- Mutable Shared Data	20	IMT
11:40	WP5- CloudButton toolkit	20	IMP
12:00	Lunch	60	
13:00	WP6- Exploitation and Dissemination	10	RHT
13:10	Final recap and summary	10	Coordinator
13:20	WP1- Project Management Use of resources, financial information Management procedures	10	Coordinator
13:30	Final remarks	10	PO
13:40	Private meeting: reviewers and PO	40	PO
14:20	Oral feedback from reviewers and comment from project	10	PO
14:30	End of Review		

*\*We assume reviewers watched videos previously*



CloudButton



Imperial College  
London



Atos



THANK YOU!



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