Serverless Data Analytics Platform

Results in Brief

Democratising big data with a new platform for cloud computing

Developing code for cloud-based big data analytics applications is complex and costly, so CloudButton has created a serverless platform of wide-ranging ability, which allocates computing resources according to demand.

DIGITAL ECONOMY

© Gorodenkoff/Shutterstock.com

Serverless data analytics platforms enable users to perform big data processing, without expert knowledge of cloud programming. They are scalable, offering massive computational and storage resources for parallel processing of terabytes of data, unlike the limited capacity of high-performance computing ↗ (HPC) clusters. They are also pay-per-use, with users only paying for resources used, billed in milliseconds, without the need for the expert IT support required by HPC clusters.

"Serverless technology can essentially democratise big data analytics; anyone with a laptop and Wi-Fi connection can take advantage of almost infinite computing resources," explains project coordinator of CloudButton ↗ Pedro Garcia Lopez, from the University of Rovira i Virgili ↗ (URV), the project host.

CloudButton created Lithops, a platform running on the same unmodified code across different cloud providers, avoiding users being locked into one vendor.

Project partner IBM is already commercialising Lithops with their clients, and Lithops

1 of 5

will be adopted by two biotech project spin-offs.

Already in incubation is the European Molecular Biology Laboratory's ↗ (EMBL) SpaceM for drug discovery and URV's DATOMA Cloud (set for 2023) which will offer cloud-based computing services for omics ↗ data.

A tool for testing times

The CloudButton team demonstrated the potential of Lithops with huge data volumes from three sources: genomics, metabolomics and geospatial. The genomics data involved compressed text, while metabolomics (the study of molecules) and geospatial data comprised large images.

For the metabolomics work, at the EMBL, a cloud-based platform called METASPACE (a previous EU project) was moved to run on top of Lithops.

"We demonstrated terabytes of metabolomics data efficiently processed in a production environment, accessed by hundreds of users around the world, including staff of organisations like AstraZeneca," adds Lopez.

Working with project partner the James Hutton Institute ↗, the team demonstrated that Lithops could improve processing performance of genomics data at reduced cost, compared to running the same analytics using an HPC cluster.

"We ran an analytical procedure called Variant Calling ↗, with a large data set on both Lithops and Illumina, a commercial option. We significantly outperformed Illumina by 3 minutes to their 30 minutes," says Lopez.

Lithops displayed the same advantages when processing geospatial data, compared to running the same code in an HPC environment.

Lithops offers support to a wide range of genomic, metabolomics and geospatial data types. It also offers a MapReduce framework, optimised to parallel process big data.

To make the system more widely available, the team have developed the CloudButton toolkit ↗, a set of open-source resources to help users migrate their applications to the Cloud in different programming languages such as Python, Java or C++.

Plug and play growth

Cloud computing is a key part of the EU's digitalisation strategy and will impact many everyday applications. CloudButton's approach could help cost-effectively simplify

this transition.

"SMEs or researchers that can't afford their own clusters or cloud experts to run code, can cheaply exploit thousands of parallel computers analysing gigabytes of data. With our system effectively hiding the back-end distributed network, users just plug and play," concludes Lopez.

The benefits are likely to be especially felt by domains like biotech and agrotech. CloudButton's toolkit could help biotech companies design new medicines, while agrotech start-ups could benefit from geospatial analytics from Sentinel 2 satellite data, for water management, for example.

Lithops will also be a key technology in three forthcoming EU-funded projects: NEARDATA (extreme data omics), CloudSkin (edge computing) and EXTRACT (extreme geospatial data), ensuring its continued development.

Keywords

CloudButton   big data   cloud computing   genomics   geospatial

metabolomics   code   programming   HPC

Discover other articles in the same domain of application

Giving a voice to voice privacy

8 April 2022

Making Europe's critical infrastructures future-proof

25 September 2024

Robot assistants help the elderly stay independent

24 August 2020

Buildings get smarter with data-driven architecture

21 February 2024

Project Information

CloudButton

Grant agreement ID: 825184

Project website ↗

DOI
10.3030/825184 ↗

Project closed

Funded under
INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Total cost
€ 4 277 507,50

EU contribution
€ 4 277 507,50

Coordinated by

EC signature date
12 November 2018

Start date
1 January 2019

End date
30 June 2022

UNIVERSITAT ROVIRA I VIRGILI
🇪🇸 Spain
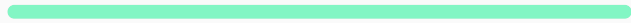
Last update: 4 November 2022

Permalink: https://cordis.europa.eu/article/id/442446-democratising-big-data-with-a-new-platform-for-cloud-computing

European Union, 2025