CloudButton

HORIZON 2020 FRAMEWORK PROGRAMME

CloudButton

(grant agreement No 825184)

Serverless Data Analytics Platform

D6.3 Final dissemination, exploitation, and adoption report

Due date of deliverable: 30-06-2022
Actual submission date: 19-07-2022

Start date of project: 01-01-2019 Duration: 36 months

Summary of the document

Document Type	Report
Dissemination level	Public
State	v1.0
Number of pages	35
WP/Task related to this document	WP6 / T6.1, T6.2
WP/Task responsible	RHAT
Leader	Tristan Tarrant (RHAT)
Technical Manager	Marc Sánchez (URV)
Quality Manager	Gerard París (URV)
Author(s)	Tristan Tarrant (RHAT), Gerard París (URV), Lara López (ATOS), Gil Vernik (IBM), Peter Pietzuch (IMP).
Partner(s) Contributing	RHAT, URV, ATOS, ANSWARE, IBM, IMP.
Document ID	CloudButton_D6.2_Public.pdf
Abstract	This document summarizes the communication and dissemination activities undertaken by the CloudButton project partners between July 2019 and June 2020.
Keywords	Communication, Dissemination, Conferences, Papers

History of changes

Version	Date	Author	Summary of changes
0.1	20-06-2020	Tristan Tarrant (RHAT)	First version.
0.2	22-07-2020	Gerard París (URV)	Reviewed. Added photos of dissemination events.
1.0	04-08-2020	Tristan Tarrant (RHAT)	Final version.
1.1	29-06-2022	Pierre Sutra (IMT)	Added contributions.
1.2	01-07-2022	Peter Pietzuch (IMP)	Added contributions.

Table of Contents

List of Abbreviations and Acronyms

ACM	Association for Computing Machinery
ASMS	American Society for Mass Spectrometry
BDV	Big Data Value
EBDVF	European Big Data Value Forum
EMBL	European Molecular Biology Laboratory
EMBO	European Molecular Biology Organization
FaaS	Function as a Service
LiDAR	Light detection and ranging
MS	Mass Spectrometry
NDVI	Normalized difference vegetation index
RHAT	Red Hat
SaaS	Software as a Service
SLA	Service-level agreement
SME	Small and medium-sized enterprises
TPI	The Pirbright Institute
URV	Universitat Rovira i Virgili
WP	Work Package

1 Executive summary

This document summarizes the communication and dissemination activities undertaken by the CloudButton project partners during the second and third semesters of the project, between July 2019 and June 2020. It also includes initial exploitation plans by industrial partners.

2 Second-year communication activities

The CloudButton partners have been active on several communication fronts over the last year, promoting the project scope and results in different ways. The following is a list of activities that have taken place since the previous report:

2.1 Online presence/Social Media/Press releases

- Promotional Video https://www.youtube.com/watch?v=y3HyDh9N8j0
- Lithops: a multi-cloud serverless programming framework https://itnext.io/lithops-a-multi-cloud-serverless-programming-framework-fd97f0d5e9e4
- Speed-up your Python applications using Lithops and serverless cloud resources https://itnext.io/speed-up-your-python-applications-using-lithops-and-serverless-cloud-resources-a64beb008bb5
- Serverless without cosnstraints https://www.ibm.com/cloud/blog/serverless-without-constraints
- Decoding dark molecular matter with IBM Cloud Functionshttps://www.ibm.com/cloud/blog/decoding-dark-molecular-matter-in-spatial-metabolomics-with-ibm-cloud-functions
- Lithops website https://lithops-cloud.github.io/
- Using Serverless to Run Your Python Code on 1000 Cores by Changing Two Lines of Code Lithops, a Multi-cloud Serverless Programming Framework https://www.ibm.com/cloud/blog/using-serverless-to-run-your-python-code-on-1000-cores-by-changing-two-lines-of-code
- Simplify the developer experience with OpenShift for Big Data processing by using Lithops framework https://medium.com/@gvernik/simplify-the-developer-experience-with-openshift-for-big-data-processing-by-using-lithops-framework-d62a795b5e1c
- CloudButton Project Overview https://www.youtube.com/watch?v=Ku9AsQlF4d8
- Geospatial Use Case - Part 1 / 2 https://www.youtube.com/watch?v=TrD2jPA-znk / https://www.youtube.com/watch?v=7K80PuHwPZ8 Serverless Big Data Analytics with Lithops - WoSC7 https://www.youtube.com/watch?v=5lCRuY82OQw
- Dissemination of CloudButton and collaboration activities on Twitter (https://twitter.com/metaspace2020)
- 06/02/2019 Project presentation post on The Matrix Foundation Website: https://bit.ly/2XupqYt
- Press release for PLANETIC (Spanish platform to foster the adoption of ICT) media channels (newsletter and website). Elaboration of a press release to advertise the project and first results within the Spanish R&D community using PLANETIC (Spanish platform for ICT promotion and adoption) media.
- Press release reporting final project achievements to be published within ATOS media channels (EN) and National media (SP).
- CloudButton article in the BDVA Newsletter (August 2021).
- Multiple blog posts on the Infinispan blog https://infinispan.org/blog/tag/cloudbutton/ related to the operator, the anchored keys, the cloud images and the native improvements
- September 19: P. Sutra, CloudButton: le Big Data à portée de clic. I'MTech, online blog. https://blogrecherche.wp.imt.fr/2019/10/10/cloudbutton-big-data-a-portee-de-clic

- Presentation of the GeoSpatial use case through the official YouTube channel of the Fundación Matrix. “Proyecto CloudButton: Caso de uso geoespacial”. `https://youtu.be/KKo8IEBEX0g`
- Presentation of CloudButton in Fundación Matrix website. Sharing main objectives and use cases. “PROGRAMA ‘HORIZON 2020’: La Fundación Matrix participa con aplicaciones para Big Data de tipo geoespacial en el proyecto CloudButton”. `https://fundacionmatrix.es/la-fundacion-matrix-participa-con-aplicaciones-para-big-data-de-tipo-geoespacial-en-el-proyecto-cloudbutton/`

2.2 Conferences, Meetings and Workshops

- P. Sutra, On the correctness of Egalitarian Paxos, Workshop on Verification (w. NETYS'19), invited talk, July 2019. http://goto.ucsd.edu/~gleissen/vds-test

  In this talk, we present a problem in both the TLA + specification and the implementation of the Egalitarian Paxos protocol. This problem is related to how replicas switch from one ballot to another when computing the dependencies of a command. We show that it may lead replicas to diverge and break the linearizability of the replicated service.

- Pedro Garcia Lopez, "CloudButton: Serverless Data Analytics". CloudButton presentation in IBM Watson Research. Open by videoconference to all IBM locations. 11 July 2019.

- BigMedilytics event (Valencia, Spain). Josep Sampé presented CloudButton project in the event "Big Data: Fueling the transformation of Europe's Healthcare Sector". 5 September 2019.

- EMBO Workshop Lipid Function in Health and Disease, Sep 2019, Dresden, DE

  Spatial metabolomics and lipidomics in tissues and single cells

- EMBL Partnership Conference Perspectives in Translational Medicine, Sep 2019, Barcelona, ES

- German Pharmaceutical Society Symposium, Sep 2019, Heidelberg, DE (keynote)

- M4I Workshop on Mass Spectrometry Imaging, Sep 2019, Maastricht, NL

- Talk at Strata Data Conference – "Your easy move to serverless computing and radically simplified data processing". https://conferences.oreilly.com/strata/strata-ny/public/schedule/detail/77226 Strata NY, September 2019

  Abstract Suppose you wrote Python code for Monte Carlo simulations to analyze financial data. The general process involves writing the code and running a simulation over small set of data to test it. Assuming this all goes smoothly, you now must run the same code at a massive scale, with parallelism, on terabytes of data, doing millions of Monte Carlo simulations. Clearly you'd prefer not to need to learn the intricacies of setting up virtual machines, suffer long setup times for the virtual machines, nor become an expert in scaling up Python code. This is exactly where serverless computing could come to the rescue. With serverless computing, you don't need to set up the computing environment and only pay for the actual amount of resources your application consumes rather than prepurchased units of capacity. Here you'll learn how to easily gain these benefits. Gil Vernik takes a deep dive into the challenge of how serverless computing can be easily used for a broad range of scenarios, like high-performance computing (HPC), Monte Carlo simulations, and data preprocessing for AI. You'll focus on how to connect existing code and frameworks to serverless without the painful process of starting from scratch and or learning new skills. To achieve this, you're based on the open source PyWren framework that introduces serverless computing with minimal effort, and its new fusion with serverless computing brings automated scalability and the use of existing frameworks into the picture. You can simply write a Python function and provide an input pointing to the dataset in a storage bucket. Then PyWren does the magic by automatically scaling and executing the user function as a serverless action at massive scale. Gil demonstrates how this capability allowed IBM to run broad range of scenarios over serverless, including Monte Carlo simulations to predict future stock prices and hyperparameter optimizations for ML models. IBM managed to complete the entire Monte Carlo simulation for stock price prediction in about 90 seconds with 1,000 concurrent invocations, compared to 247 minutes with almost 100% CPU utilization running the same flow over a laptop with 4 CPU cores. He'll also show you how to bond TensorFlow and serverless for the data-preparation phases. Existing TensorFlow code can be easily adapted and benefit serverless with only minimal code modifications and without users having to learn serverless architectures and deployments.

- ASMS Asilomar Conference on Mass Spectrometry, Oct 2019, Pacific Grove, CA, USA
  Spatial Metabolomics: From Big Data to Single Cells,
- Munich Metabolomics Meeting, Oct 2019, Munich, DE (keynote)
- Bayer MS Imaging Workshop, Oct 2019, Berlin, DE
- OurCon conference on Imaging Mass Spectrometry, Oct 2019, St. Malo, France
- Talk at EBDVF in Helsinki, "Your easy move to serverless computing and radically simplified data processing", `https://www.european-big-data-value-forum.eu/program-day-2/` European Big Data Value Forum, October 2019
- Talk at Haifa tech meetup – "Your easy move to serverless computing". `https://www.meetup.com/Haifa-Tech-Talks/events/261393733/`
- Presentation to Intel Labs of the current status of the Faasm runtime, October 2019.
- Presentation of CloudButton as one of the most prominent ATOS projects within the online boot at EBDVF, November 2021.
- Presentation of Faasm to Michael Berendt by Peter Pietzuch and Simon Shillaker (IMP), IBM Distinguished Engineer/CTO Serverless, November 2019.
- Pedro Garcia Lopez, Talk: "The next revolution in Distributed Computing". IBM Watson Research. 21 January 2020.
- Presentation of Faasm/CloudButton by Peter Pietzuch (IMP) to Intel Labs, Portland, Oregon, USA, May 2020.
- "Serverless Computing Panel", Presentation of CloudButton/Faasm by Peter Pietzuch (IMP), University of Waterloo, August 2020.
- Workshop "Next-Generation Cloud Technologies", Presentation by Peter Pietzuch (IMP) of CloudButton/Faasm, organised by Huawei Munich Research Centre, November 2020.
- Talk on "Faasm/WebAssembly for Serverless" by Simon Shillaker (IMP) at 2nd Workshop on Resource Disaggregation and Serverless (WORDS'21), co-located with ASPLOS 2021, April 2021.
- Pierre Sutra (IMT) participated to Sciences Num. (online podcast), May 2021.
- Invitation for Peter Pietzuch (IMP) to talk about Faasm/CloudButton at Microsoft Research Summit 2021, "The Future of Cloud is Serverless", October 2021.
- Invited talk by Peter Pietzuch (IMP) on CloudButton/Faasm, Huawei Strategy & Technology Workshop "Heterogeneous System Software", October 2021.
- Anatole Lefort (IMT) participated to NVMW'22, UC San Diego, May 2022.
- Anatole Lefort (IMT) gave an invited talk at HMEM'22, online, June 2022.
- Pierre Sutra (IMT) gave an invited talk at PER3S'22, Palaiseau, June 2022.
- Alexander Theodorov (EMBL) KubeCon, May 2021, online
- Alexander Theodorov (EMBL) IBM Cloud Native Day, Apr 2021, online
- Alexander Theodorov (EMBL) Nature Conference Bioengineering Solutions for Biology & Medicine (BIOENG), Jul 2022, Munich, DE

- Alexander Theodorov (EMBL) Single-Cell Proteomics Conference (SCP2022), Jun 2022, Boston, MA, USA (virtual)
- Alexander Theodorov (EMBL) Int. Conference of TRR179 Viral Hepatitis and Beyond, May 2022, Freiburg, DE
- Alexander Theodorov (EMBL) Spatial Omics SEMM Roundtable at the European School of Molecular Medicine, May 2022, Milan, IT (virtual)
- Alexander Theodorov (EMBL) EMBL-EBI Training Course "Introduction to Metabolomics Analysis", May 2022, Hinxton, UK
- Alexander Theodorov (EMBL) Mitteldeutsche Laborkonferenz (Central Germany Diagnostics Lab Conference), May 2022, Weimar, DE
- Alexander Theodorov (EMBL) 2nd Annual Mini Symposium in Human Genetics, Apr 2022, Emory Uni / Atlanta, GA, USA
- Alexander Theodorov (EMBL) Spatial Biology UK, Apr 2022, London, UK
- Alexander Theodorov (EMBL) Science in Immunology Meeting at King's College London, Apr
- Alexander Theodorov (EMBL) 2022, London, UK (keynote)
- Alexander Theodorov (EMBL) Chan-Zuckerberg Initiative (CZI) workshop on single-cell proteomics and metabolomics, Apr 2022, US / virtual
- Alexander Theodorov (EMBL) Spatial Biology – Accelerating Translational Research & Clinical Applications, Apr 2022, UK / virtual
- Alexander Theodorov (EMBL) Korean Metabolomics Society, Apr 2022, South Korea / virtual (keynote)
- Alexander Theodorov (EMBL) Online Round Table of Institute of Biomedicine of Seville, Mar 2022, Sevilla, ES / virtual
- Alexander Theodorov (EMBL) MINDS Workshop, Mar 2022, Wiesbaden, DE
- Alexander Theodorov (EMBL) Certificate of Advanced Studies In Personalized Molecular Oncology, Feb 2022, Basel, CH / virtual
- Alexander Theodorov (EMBL) German Mass Cytometry User Forum, Jan 2022, Berlin, DE / online
- Alexander Theodorov (EMBL) Int. Conference of SFB1335 Aberrant Immune Signals in Cancer, Dec 2021, Munich, DE / online
- Alexander Theodorov (EMBL) Pioneers in Omics, Dec 2021, virtual
- Alexander Theodorov (EMBL) BigData World Congress and Genomics Live, Nov 2021, Basel, CH
- Alexander Theodorov (EMBL) Int. Mass Spectrometry Imaging Conference (OurCon), Oct 2021, UK / virtual (keynote)
- Alexander Theodorov (EMBL) CSHL meeting on Mechanisms of Metabolic Signaling, Oct 2021, Cold Spring Harbor, NY, USA
- Alexander Theodorov (EMBL) Oz Single-Cell Brisbane Meeting, Oct 2021, Brisbane, AU / online

- Alexander Theodorov (EMBL) Plant Cell Atlas Symposium, Oct 2021, online
- Alexander Theodorov (EMBL) de.NBI Summer School Mass Spec Omics Data, Sep 2021, online
- Alexander Theodorov (EMBL) 2nd Workshop on Metabolomics in Chronic Inflammation Barrier Diseases, Sep 2021, Kiel, DE
- Alexander Theodorov (EMBL) ISMB, CompMS session, July 2021, online (keynote)
- Alexander Theodorov (EMBL) ICI-CIVIS Summer School in Immuno-Oncology, July 2021, Marseille, FR (keynote)
- Alexander Theodorov (EMBL) Metabolomics Conference, June 2021, online (keynote)
- Alexander Theodorov (EMBL) IHMC International Human Microbiome Consortium Congress, Jun 2021, ES / online (keynote)
- Alexander Theodorov (EMBL) LIPID MAPS Spring School, Apr 2021, online
- Alexander Theodorov (EMBL) Keystone Symposium on Metabolic Decisions in Development & Disease, Mar 2021, US / online
- Alexander Theodorov (EMBL) VIZBI Visualizing Biological Data Workshop, Mar 2021, DE / online
- Alexander Theodorov (EMBL) 72nd AACC Annual Scientific Meeting, Dec 2020, US / online
- Alexander Theodorov (EMBL) START-HD Single-Cell and Spatial Omics Workshop, Nov 2020, Heidelberg / online
- Alexander Theodorov (EMBL) Lipid Droplets Workshop, Nov 2020, online
- Alexander Theodorov (EMBL) BigData World Congress and Genomics Live, Nov 2020, Basel, CH / online
- Alexander Theodorov (EMBL) FASTest 2020 E-lecture for Dutch Tenure Track Network & COAST, Sep 2020, online
- Alexander Theodorov (EMBL) Spatial Omics, Joint Zoom Seminar Series http://spatialomics.net, Sep 2020, online
- Alexander Theodorov (EMBL) 12th Annual Congress in Clinical Mass Spectrometry (MSACL), May 2020, online (keynote)
- Alexander Theodorov (EMBL) Webinar in Sphingolipids Biology, May 2020, online
- Alexander Theodorov (EMBL) Certificate of Advanced Studies In Personalized Molecular Oncology, Feb 2020, Basel, CH
- Alexander Theodorov (EMBL) 2nd Annual Symposium of ITMO on Immunology, Inflammation, Infectiology and Microbiology (I3M), Feb 2020, Lille, FR (keynote)
- Alexander Theodorov (EMBL) EMBO-EMBL Symposium Metabolism Meets Epigenetics, Nov 2019, Heidelberg, DE
- Alexander Theodorov (EMBL) ASMS Asilomar Conference on Mass Spectrometry, Oct 2019, Pacific Grove, CA, USA
- Alexander Theodorov (EMBL) Munich Metabolomics Meeting, Oct 2019, Munich, DE (keynote)

- Alexander Theodorov (EMBL) Bayer MS Imaging Workshop, Oct 2019, Berlin, DE
- Alexander Theodorov (EMBL) EMBO Workshop Lipid Function in Health and Disease, Sep 2019, Dresden, DE
- Alexander Theodorov (EMBL) EMBL Partnership Conference Perspectives in Translational Medicine, Sep 2019, Barcelona, ES
- Alexander Theodorov (EMBL) German Pharmaceutical Society Symposium, Sep 2019, Heidelberg, DE (keynote)
- Alexander Theodorov (EMBL) M4I Workshop on Mass Spectrometry Imaging, Sep 2019, Maastricht, NL
- Alexander Theodorov (EMBL) Metabolomics Conference, Jun 2019, the Hague, NL (keynote)
- Alexander Theodorov (EMBL) BMBF e:Med Summer School, May 2019, Castle Ringberg, DE
- Alexander Theodorov (EMBL) Highlight Symposium at European Molecular Imaging Meeting, March 2019, Glasgow, UK
- Alexander Theodorov (EMBL) VIB Conference Metabolism in Cancer and Stromal Cells, Nov 2018, Leuven, BE (plenary)
- Alexander Theodorov (EMBL) Clinical Proteomics Symposium, Nov 2018, Barcelona, ES
- Alexander Theodorov (EMBL) OurCon & IMSS conference on Imaging Mass Spectrometry, Nov 2018, Charleston, NC, USA
- Alexander Theodorov (EMBL) Multi-Omics Seminar Tour by ThermoFisher Scientific, Oct 2018, Wiesloch, DE
- Alexander Theodorov (EMBL) AI and Blockchain for Healthcare Forum at Based Life Conference, Sep 2018, Basel, CH
- Alexander Theodorov (EMBL) Helmholtz Conference on Engineering Biomedical & Metabolic Breakthroughs, Jul 2018, Venice, IT
- Articles about the impact of Lithops on many national newspapers and media: Europa Press, La Vanguardia, InfoCamp, Diari Més, Tarragona Ràdio, Reus Digital, Gente digital, Nació digital, Cope, Cadena Ser, 5 Jul 2022
- Pedro Garcia presents CloudButton results in the Huawei InnoWave Global Conference
- 16 Jun 2022: Pedro Garcia Lopez chairs WOSCx1, 21 Jun 2022 https://www.serverlesscomputing.org/wosc (first online Workshop on Serverless Computing). In the panel, CloudButton contributions to the field, including Lithops, are discussed
- 3 Jun 2022: Pedro Garcia presents CloudButton in the panel "Using Big Data and Data Marketplaces to power Digital Transformation for European Ecosystem and SMEs" in the context of DataWeek'22 https://www.big-data-value.eu/data-week-2022/
- 18 May 2022: Marc Sanchez and Raul Saiz-Laudo speak at CCGrid symposium, and present "Egeon: Software-Defined Data Protection for Object Storage" https://fcrlab.unime.it/ccgrid22/wp-content/uploads/2022/04/CCGRID22-AdvancedProgram.pdf
- 12 May 2022: The DATOMA project, a startup on OMICS Data Analytics emerged from CloudButton and YanesLab, is presented in the second round of La Caixa Research Validate

- 15 Feb 2022: Pedro Garcia presents Cloudbutton in the HIPEAC workshop
- Dec 2021: URV presents "Serverless Big Data Analytics with Lithops" in the Serverless Data Analytics in the WOSC7 Workshop (International Workshop on Serverless Computing) in the session for Serverless Data Analytics organized by CloudButton
- 29 Nov 2021: Pedro Garcia, Josep Sampe, Gil Vernik, Tristan Tarrant present CloudButton in the European Big Data Value Forum 2021
- 18 Oct 2021: Pedro Garcia presents "Data-driven Serverless Analytics over Cloud Object Storage" in the Huawei Strategy and Technology Workshop 2021 (STW 21) 2-3 Sep 2021: Presentation at JCSD 2021 (Jornadas de Concurrencia y Sistemas Distribuidos): interventions by URV about "Benchmarking parallelism in FaaS platforms", Primula, and Triggerflow
- Jul-Sep 2021: Horizon Results Booster (With EMBL). Invitation to Huawei STW (Strategy and Technology Workshop) about Data Driven Serverless Analytics.
- Apr-Jun 2021: Meeting with Albert Tenesa. University of Edinburgh. Cloud and Serverless Genomics. Interested party in the CloudButton project. Meeting with Beril Kirci. PMO Partners. Interest from Doğuş Technology in CloudButton results. The company aims to become a Stakeholder of CloudButton.eu. Meeting with IBM Spain (Oriol Vilaplana Canal). Cloudbutton dissemination.
- 14 May 2021: Pedro Garcia is the keynote speaker at the Serverless Computing Seminar at Schloss Dagstuhl https://www.dagstuhl.de/en/programm/kalender/semhp/?semnr=21201
- 25 Nov 2021: CloudButton is present at the European Researchers Night by ComCiència (URV) http://tarragona.nitdelarecerca.cat/ca/european-corner/
- 22 Mar 2021: CloudButton announces its first challenge open to Computer Engineering Students https://challenge.cloudbutton.eu/ to analyze big datasets using the Lithops toolkit and Cloud distributed systems. URV promotes participation from its students
- 8 Dec 2020: Participation in International Workshop on Serverless Computing WOSC6, Mariano Mirabelli presents "Bringing scaling transparency to Proteomics applications with serverless computing" by URV https://www.youtube.com/watch?v=4GSQRNgdF0A
- 21 Jan 2020: Pedro Garcia talk: "The next revolution in Distributed Computing" in IBM Watson Research
- 19 Dec 2019: Daniel Barcelona presents "On the FaaS track: Building Stateful Distributed Applications with Serverless Architectures" at the Middleware 2019 conference, UC Davis, California
- 5 Sep 2019: Josep Sampe presents CloudButton at the Big Medilytics Event 2019 in Valencia, Spain
- 16 Jul 2019: Pedro Garcia presents "Please, do not Decentralize the Internet with (Permissionless) Blockchains!" at the 39th IEEE International Conference on Distributed Computing Systems (ICDCS) in Dallas, Texas
- 11 Jul 2019: Pedro Garcia presents "CloudButton: Serverless Data Analytics" in IBM Watson Research, open by videoconference to all IBM locations
- 26-28 Jun 2019: URV attendance at Big Data Value PPP Summit, Riga
- 13-15 Jun 2019: Pedro Garcia attends JSCD 2019 (XXV Jornadas de Concurrencia y Sistemas Distribuidos) in Zaragoza, Spain, to disseminate preliminary results of the CloudButton project

- 20 May 2019: Marc Sanchez presents "LambdaFlow: Automatic Pushdown of Dataflow Operators Close to the Data" in the 19th IEEE/ACM Cluster, Cloud and Grid Computing (CCGrid) Symposium 2019, Larnaca, Cyprus http://www.ccgrid2019.org/
- 27 Feb 2019: BDVA meeting in Brussels: Pedro Garcia presents CloudButton to BDVA community. Interactions with other H2020 projects
- Feb-Mar 2019: CloudButton announcement in Tarragona: URV press announces CloudButton https://diaridigital.urv.cat/la-urv-coordina-un-projecte-europeu-per-simplificar-lanalisi-massiva-de-dades-al-nuvol/
- Pedro Garcia speaks about CloudButton in Onda Cero radio
- Pedro Garcia speaks about CloudButton in Tarragona Radio
- Pedro Garcia speaks about CloudButton and Serverless to the Tarragona Developers community
- Interview at medium.com about Serverless Orchestration Services
- 29 Sep 2018: CloudButton concept is present at the European Researchers Night in Reus (talk by Pedro Garcia)

2.3 Conference papers, journal articles and preprints

- Raúl Gracia Tinedo, Marc Sánchez Artigas, Pedro García López, Yosef Moatti, Filip Gluszak: Lambda-Flow: Automatic Pushdown of Dataflow Operators Close to the Data. 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID 2019), Larnaca, Cyprus, May 14-17, 2019. pp. 112-121.

  Abstract Modern data analytics infrastructures are composed of physically disaggregated compute and storage clusters. Thus, dataflow analytics engines, such as Apache Spark or Flink, are left with no choice but to transfer datasets to the compute cluster prior to their actual processing. For large data volumes, this becomes problematic, since it involves massive data transfers that exhaust network bandwidth, that waste compute cluster memory, and that may become a performance barrier. To overcome this problem, we present λFlow: a framework for automatically pushing dataflow operators (e.g., map, flatMap, filter, etc.) down onto the storage layer. The novelty of λFlow is that it manages the pushdown granularity at the operator level, which makes it a unique problem. To wit, it requires addressing several challenges, such as how to encapsulate dataflow operators and execute them on the storage cluster, and how to keep track of dependencies such that operators can be pushed down safely onto the storage layer. Our evaluation reports significant reductions in resource usage for a large variety of IO-bound jobs. For instance, λFlow was able to reduce both network bandwidth and memory requirements by 90% in Spark. Our Flink experiments also prove the extensibility of λFlow to other engines.

- Pedro García López, Alberto Montresor, Anwitaman Datta: Please, do not Decentralize the Internet with (Permissionless) Blockchains!. 39th IEEE International Conference on Distributed Computing Systems, ICDCS 2019, Dallas, TX, USA, July 7-10, 2019. pp. 1901-1911.

  Abstract The old mantra of decentralizing the Internet is coming again with fanfare, this time around the blockchain technology hype. We have already seen a technology supposed to change the nature of the Internet: peer-to-peer. The reality is that peer-to-peer naming systems failed, peer-to-peer social networks failed, and yes, peer-to-peer storage failed as well. In this paper, we will review the research on distributed systems in the last few years to identify the limits of open peer-to-peer networks. We will address issues like system complexity, security and frailty, instability and performance. We will show how many of the aforementioned problems also apply to the recent breed of permissionless blockchain networks. The applicability of such systems to mature industrial applications is undermined by the same properties that make them so interesting for a libertarian audience: namely, their openness, their pseudo-anonymity and their unregulated cryptocurrencies. As such, we argue that permissionless blockchain networks are unsuitable to be the substrate for a decentralized Internet. Yet, there is still hope for more decentralization, albeit in a form somewhat limited with respect to the libertarian view of decentralized Internet: in cooperation rather than in competition with the superpowerful datacenters that dominate the world today. This is derived from the recent surge in interest in byzantine fault tolerance and permissioned blockchains, which opens the door to a world where use of trusted third parties is not the only way to arbitrate an ensemble of entities. The ability of establish trust through permissioned blockchains enables to move the control from the datacenters to the edge, truly realizing the promises of edge-centric computing.

- Pedro García-López: Decentralize the feedback infrastructure!. CoRR abs/2010.03356

  Abstract The decentralized architecture of Internet sparkled techno-utopian visions of a virtual freedom space for humanity. Peer-to-peer systems, collaborative creation (wikipedia), open source software (Linux), universal shared knowledge, and the hopes for disintermediation contributed to this major vision. However, the reality is bleak: centralization is reigning in the cyberspace, with huge technological corporations controlling our data, and re-intermediation and control are stronger than ever in the so-called "sharing" economy. The Internet is also fragmented by countries, with many states imposing heavy controls to information and communication services. The XXI century will witness the major clash between centralization and decentralization in human history. And the major struggle will be around the communication and feedback technologies that will intermediate and govern every interaction in our lives. Unlike previous approaches that propose to socialize the feedback infrastructure or to use anti-monopoly laws to break Big Tech companies, in this article we advocate for the decentralization of the information and communication infrastructure. And the key to this decentralization is the creation of standards enabling interoperability between data platforms. This will in turn produce a true disintermediation from well established technological players and open competition to small third parties. In this article, we sketch such a decentralized open infrastructure including communication, sharing, matchmaking, and reputation services that can be constructed over open source technologies and standards.

- Pedro García López, Marc Sánchez Artigas, Simon Shillaker, Peter R. Pietzuch, David Breitgand, Gil Vernik, Pierre Sutra, Tristan Tarrant, Ana Juan Ferrer: ServerMix: Tradeoffs and Challenges of Serverless Data Analytics. arXiv CoRR abs/1907.11465 (July 2019).

  Abstract Serverless computing has become very popular today since it largely simplifies cloud programming. Developers do not need to longer worry about provisioning or operating servers, and they pay only for the compute resources used when their code is run. This new cloud paradigm suits well for many applications, and researchers have already begun investigating the feasibility of serverless computing for data analytics. Unfortunately, today's serverless computing presents important limitations that make it really difficult to support all sorts of analytics workloads. This paper first starts by analyzing three fundamental trade-offs of today's serverless computing model and their relationship with data analytics. It studies how by relaxing disaggregation, isolation, and simple scheduling, it is possible to increase the overall computing performance, but at the expense of essential aspects of the model such as elasticity, security, or sub-second activations, respectively. The consequence of these trade-offs is that analytics applications may well end up embracing hybrid systems composed of serverless and serverful components, which we call Servermix in this paper. We will review the existing related work to show that most applications can be actually categorized as Servermix. Finally, this paper will introduce the major challenges of the CloudButton research project to manage these trade-offs.

- Mariano Ezequiel Mirabelli, Pedro García-López, Gil Vernik: Bringing scaling transparency to Proteomics applications with serverless computing WoSC '20: 6th International Workshop on Serverless Computing

  Abstract Scaling transparency means that applications can expand in scale without changes to the system structure or the application algorithms. Serverless Computing's inherent auto-scaling support and fast function launching is ideally suited to support scaling transparency in different domains. In particular, Proteomic applications could considerably benefit from scaling transparency and serverless technologies due to their high concurrency requirements. Therefore, the auto-provisioning nature of serverless platforms makes this computing model an alternative to satisfy dynamically the resources required by protein folding simulation processes. However, the transition to these architectures must face challenges: they should show comparable performance and cost to code running in Virtual Machines (VMs). In this article, we demonstrate that Proteomics applications implemented with the Replica Exchange algorithm can be moved to serverless settings guaranteeing scaling transparency. We also validate that we can reduce the total execution time by around forty percent with comparable cost to cluster technologies (Work Queue) over VMs.

- Marc Sánchez-Artigas, Germán T. Eizaguirre, Gil Vernik, Lachlan Stuart, Pedro García-López: Primula: a Practical Shuffle/Sort Operator for Serverless Computing Middleware '20: 21st International Middleware Conference Industrial Track **Abstract** Serverless computing has recently gained much attention as a feasible alternative to always-on IaaS for data processing.

However, existing severless frameworks are not (yet) usable enough to reach out to a large number of users. To wit, they still require developers to specify the number of serverless functions for a simple sort job. We report our experience in designing Primula, a serverless sort operator that abstracts away users from the complexities of resource provisioning, skewed data and stragglers, yielding the most accessible sort primitive to date. Our evaluation on the IBM Cloud platform demonstrates the usability of Primula without abandoning performance (e.g., 3x faster than a serverless Spark backend and 62% slower than a hybrid serverless/IaaS solution).

- Katja Ovchinnikova, Alexander Rakhlin, Lachlan Stuart, Sergey Nikolenko, Theodore Alexandrov, ColocAI: artificial intelligence approach to quantify co-localization between mass spectrometry images, BioRxiv, `https://www.biorxiv.org/content/10.1101/758425v1`, September 2019.

  Abstract Motivation Imaging mass spectrometry (imaging MS) is a prominent technique for capturing distributions of molecules in tissue sections. Various computational methods for imaging MS rely on quantifying spatial correlations between ion images, referred to as co-localization. However, no comprehensive evaluation of co-localization measures has ever been performed; this leads to arbitrary choices and hinders method development.

  Results We present ColocAI, an artificial intelligence approach addressing this gap. With the help of 42 imaging MS experts from 9 labs, we created a gold standard of 2210 pairs of ion images ranked by their co-localization. We evaluated existing co-localization measures and developed novel measures using tf-idf and deep neural networks. The semi-supervised deep learning Pi model and the cosine score applied after median thresholding performed the best (Spearman 0.797 and 0.794 with expert rankings respectively). We illustrate these measures by inferring co-localization properties of 10273 molecules from 3685 public METASPACE datasets.

- Alexey Gotsman, Anatole Lefort, Gregory V. Chockler. White-Box Atomic Multicast. 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2019, Portland, OR, USA, June 24-27, 2019,. pp. 176-87.

  Abstract Atomic multicast is a communication primitive that delivers messages to multiple groups of processes according to some total order, with each group receiving the projection of the total order onto messages addressed to it. To be scalable, atomic multicast needs to be genuine, meaning that only the destination processes of a message should participate in ordering it. In this paper we propose a novel genuine atomic multicast protocol that in the absence of failures takes as low as 3 message delays to deliver a message when no other messages are multicast concurrently to its destination groups, and 5 message delays in the presence of concurrency. This improves the latencies of both the fault-tolerant version of classical Skeen's multicast protocol (6 or 12 message delays, depending on concurrency) and its recent improvement by Coelho et al. (4 or 8 message delays). To achieve such low latencies, we depart from the typical way of guaranteeing fault-tolerance by replicating each group with Paxos. Instead, we weave Paxos and Skeen's protocol together into a single coherent protocol, exploiting opportunities for white-box optimisations. We experimentally demonstrate that the superior theoretical characteristics of our protocol are reflected in practical performance pay-offs.

- Daniel Barcelona Pons, Marc Sánchez Artigas, Gerard París, Pierre Sutra, Pedro García López: On the FaaS Track: Building Stateful Distributed Applications with Serverless Architectures. 20th International Middleware Conference, Middleware 2019, Davis, CA, USA, December 9-13, 2019. pp. 41-54.

  Abstract Serverless computing is an emerging paradigm that greatly simplifies the usage of cloud resources and suits well to many tasks. Most notably, Function-as-a-Service (FaaS) enables programmers to develop cloud applications as individual functions that can run and scale independently. Yet, due to the disaggregation of storage and compute resources in FaaS, applications that require fine-grained support for mutable state and synchronization, such as machine learning and scientific computing, are hard to build. In this work, we present Crucial, a system to program highly-concurrent stateful applications with serverless architectures. Its programming model keeps the simplicity of FaaS and allows to port effortlessly multi-threaded algorithms to this new environment. Crucial is built upon the key insight that FaaS resembles to concurrent programming at the scale of a data center. As a consequence, a distributed shared memory layer is the right answer to the need for fine-grained state management and coordination in serverless. We validate our system with the help of micro-benchmarks and various applications. In particular, we implement two common machine learning algorithms: k-means clustering and logistic regression. For both cases, Crucial obtains superior or comparable performance to an equivalent Spark cluster.

- Daniel Barcelona Pons, Pedro García López, Álvaro Ruiz Ollobarren, Amanda Gómez-Gómez, Gerard París, Marc Sánchez Artigas: FaaS Orchestration of Parallel Workloads. 5th International Workshop on Serverless Computing, WOSC@Middleware 2019, Davis, CA, USA, December 09-13, 2019. pp. 25-30.

  Abstract Function as a Service (FaaS) is based on a reactive programming model where functions are activated by triggers in response to cloud events (e.g., objects added to an object store). The inherent elasticity and the pay-per-use model of serverless functions make them very appropriate for embarrassingly parallel tasks like data preprocessing, or even the execution of MapReduce jobs in the cloud.

  But current Serverless orchestration systems are not designed for managing parallel fork-join workflows in a scalable and efficient way. We demonstrate in this paper that existing services like AWS Step Functions or Azure Durable Functions incur in considerable overheads, and only Composer at IBM Cloud provides suitable performance.

  Successively, we analyze the architecture of OpenWhisk as an open-source FaaS systems and its orchestration features (Composer). We outline its architecture problems and propose guidelines for orchestrating massively parallel workloads using serverless functions.

- Vitor Enes, Carlos Baquero, Tuanir França Rezende, Alexey Gotsman, Matthieu Perrin, Pierre Sutra: State-Machine Replication for Planet-Scale Systems. EuroSys'20: Fifteenth EuroSys Conference 2020, Heraklion, Greece, April 27-30, 2020. pp 24:1-24:15.

  Abstract Online applications now routinely replicate their data at multiple sites around the world. In this paper we present Atlas, the first state-machine replication protocol tailored for such planet-scale systems. Atlas does not rely on a distinguished leader, so clients enjoy the same quality of service independently of their geographical locations. Furthermore, client-perceived latency improves as we add sites closer to clients. To achieve this, Atlas minimizes the size of its quorums using an observation that concurrent data center failures are rare. It also processes a high percentage of accesses in a single round trip, even when these conflict. We experimentally demonstrate that Atlas consistently outperforms state-of-the-art protocols in planet-scale scenarios. In particular, Atlas is up to two times faster than Flexible Paxos with identical failure assumptions, and more than doubles the performance of Egalitarian Paxos in the YCSB benchmark.

- P. Sutra: On the Correctness of Egalitarian Paxos, Information Processing Letters, Vol. 156, April 2020.

  Abstract This paper identifies a problem in both the TLA+ specification and the implementation of the Egalitarian Paxos protocol. It is related to how replicas switch from one ballot to another when computing the dependencies of a command. The problem may lead replicas to diverge and break the linearizability of the replicated service.

- Pedro García López, Aleksander Slominski, Simon Shillaker, Michael Behrendt, Bernard Metzler: Serverless End Game: Disaggregation enabling Transparency. arXiv CoRR abs/2006.01251 (June 2020).

  Abstract For many years, the distributed systems community has struggled to smooth the transition from local to remote computing. Transparency means concealing the complexities of distributed programming like remote locations, failures or scaling. For us, full transparency implies that we can compile, debug and run unmodified single-machine code over effectively unlimited compute, storage, and memory resources. We elaborate in this article why resource disaggregation in serverless computing is the definitive catalyst to enable full transparency in the Cloud. We demonstrate with two experiments that we can achieve transparency today over disaggregated serverless resources and obtain comparable performance to local executions. We also show that locality cannot be neglected for many problems and we present five open research challenges: granular middleware and locality, memory disaggregation, virtualization, elastic programming models, and optimized deployment. If full transparency is possible, who needs explicit use of middleware if you can treat remote entities as local ones? Can we close the curtains of distributed systems complexity for the majority of users?

- Pedro García López, Aitor Arjona, Josep Sampé, Aleksander Slominski, Lionel Villard: Triggerflow: Trigger-based orchestration of Serverless Workflows. DEBS '20: The 14th ACM International Conference on Distributed and Event-based Systems, Montreal, Quebec, Canada, July 13-17, 2020. pp. 3-14.

  Abstract As more applications are being moved to the Cloud thanks to serverless computing, it is increasingly necessary to support native life cycle execution of those applications in the data center. But existing systems either focus on short-running workflows (like IBM Composer or Amazon Express Workflows) or impose considerable overheads for synchronizing massively parallel jobs (Azure Durable Functions, Amazon Step Functions, Google Cloud Composer). None of them are open systems enabling extensible interception and optimization of custom workflows. We present Triggerflow: an extensible Trigger-based Orchestration architecture for serverless workflows built on top of Knative Eventing and Kubernetes technologies. We demonstrate that Triggerflow is a novel serverless building block capable of constructing different reactive schedulers (State Machines, Directed Acyclic Graphs, Workflow as code). We also validate that it can support high-volume event processing workloads, auto-scale on demand and transparently optimize scientific workflows.

- Simon Shillaker, Peter Pietzuch: Faasm: Lightweight Isolation for Efficient Stateful Serverless Computing, USENIX Annual Technical Conference (USENIX ATC), 15-17 July 2020, Boston, MA, USA.

  Abstract Serverless computing is an excellent fit for big data processing because it can scale quickly and cheaply to thousands of parallel functions. Existing serverless platforms isolate functions in ephemeral, stateless containers, preventing them from directly sharing memory. This forces users to duplicate and serialise data repeatedly, adding unnecessary performance and resource costs. We believe that a new lightweight isolation approach is needed, which supports sharing memory directly between functions and reduces resource overheads. We introduce Faaslets, a new isolation abstraction for high-performance serverless computing. Faaslets isolate the memory of executed functions using software-fault isolation (SFI), as provided by WebAssembly, while allowing memory regions to be shared between functions in the same address space. Faaslets can thus avoid expensive data movement when functions are co-located on the same machine. Our runtime for Faaslets, Faasm, isolates other resources, e.g. CPU and network, using standard Linux cgroups, and provides a low-level POSIX host interface for networking, file system access and dynamic loading. To reduce initialisation times, Faasm restores Faaslets from already-initialised snapshots. We compare Faasm to a standard container-based platform and show that, when training a machine learning model, it achieves a 2x speed-up with 10x less memory; for serving machine learning inference, Faasm doubles the throughput and reduces tail latency by 90%.

- Gerard París, Pedro García López, Marc Sánchez-Artigas. Serverless Elastic Exploration of Unbalanced Algorithms, 13th IEEE International Conference on Cloud Computing, CLOUD 2020, October, 2020.

  Abstract In recent years, serverless computing and, in particular the Function-as-a-Service (Faas) execution model, has proven to be efficient for running parallel computing tasks. However, little attention has been paid to highly-parallel applications with unbalanced and irregular workloads. The main challenge of executing this type of algorithms in the cloud is the difficulty to account for the computing requirements beforehand. This places a burden on scientific users who very often make bad decisions by either overprovisioning resources or inadvertently limiting the parallelism of these algorithms due to resource contention. Our hypothesis is that the elasticity and ease of management of serverless computing can help users avoid such decisions, which may lead to undesirable cost-performance consequences for unbalanced problem spaces. In this work, we show that with a simple serverless executor pool abstraction, which can combine local resources with remote cloud functions, one can achieve a better cost-performance trade-off than a Spark cluster of static size and large EC2 VMs. To support this conclusion, we evaluate two unbalanced algorithms: the Unbalanced Tree Search (UTS) and the Mandelbrot Set using the Mariani-Silver algorithm. For instance, our serverless implementation of UTS is able to outperform Spark by up to 55% with the same cost. This provides the first concrete evidence that highly-parallel, irregular workloads can be efficiently executed using purely stateless functions with almost zero burden on users — i.e., no need for users to understand non-obvious system-level parameters and optimizations.

- M Wagenländer, L Mai, G Li, P Pietzuch. Spotnik: Designing distributed machine learning for transient cloud resources 12th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 2020), Boston, MA, USA.

  Abstract To achieve higher utilisation, cloud providers offer VMs with GPUs as lower-cost transient cloud resources. Transient VMs can be revoked at short notice and vary in their availability. This poses challenges to distributed machine learning (ML) jobs, which perform long-running stateful computation in which many workers maintain and synchronise model replicas. With transient VMs, existing systems either require a fixed number of reserved VMs or degrade performance when recovering from revoked transient VMs.

  We believe that future distributed ML systems must be designed from the ground up for transient cloud resources. This paper describes *SPOTNIK*, a system for training ML models that features a more adaptive design to accommodate transient VMs: (i) SPOTNIK uses an adaptive implementation of the all-reduce collective communication operation. As workers on transient VMs are revoked, SPOTNIK updates its membership and uses the all-reduce ring to recover; and (ii) SPOTNIK supports the adaptation of the synchronisation strategy between workers. This allows a training job to switch between different strategies in response to the revocation of transient VMs. Our experiments show that, after VM revocation, SPOTNIK recovers training within 300 ms for ResNet/ImageNet.

- Tuanir França Rezende, Pierre Sutra. Leaderless State-Machine Replication: Specification, Properties, Limits, 34th International Symposium on Distributed Computing, DISC 2020, October, 2020.

  Abstract Modern Internet services commonly replicate critical data across several geographical locations using state-machine replication (SMR). Due to their reliance on a leader replica, classical SMR protocols offer limited scalability and availability in this setting. To solve this problem, recent protocols follow instead a leaderless approach, in which each replica is able to make progress using a quorum of its peers. In this paper, we study this new emerging class of SMR protocols and states some of their limits. We first propose a framework and compelling set of definitions that capture the essence of leaderless state-machine replication (Leaderless SMR). Then, we introduce a set of desirable properties for these protocols: (R)eliability, (O)ptimal (L)atency and (L)oad Balancing. We show that protocols matching all of the ROLL properties are subject to a trade-off between performance and reliability. We also establish a lower bound on the message delay to execute a command in protocols optimal for the ROLL properties. This lower bound explains the persistent chaining effect observed in experimental results.

- Aurèle Mahéo, Pierre Sutra, Tristan Tarrant. The Serverless Shell, 22nd International Middleware Conference (industry track), Middleware 2021, online, 2021.

  Abstract Serverless computing is a recent paradigm to program the cloud. It allows to execute user-defined functions at scale, on demand, and in a pay-per-use manner. This paper reports on adapting the Unix shell for serverless. Our software, called the serverless shell (`sshell`), runs shell scripts on a serverless platform much like with a regular computer. It permits to reuse an existing code base while benefiting from the massive power of serverless and paying only for the resources used. `sshell`is built around a small set of components that includes a new inter-process communication layer for serverless. We evaluate it in AWS Lambda using several micro-benchmarks and a large-scale application. Our results show that `sshell`achieves comparable or better performance than a high-end server. Moreover, it can be faster and more cost-efficient than a cluster-based solution to mine large datasets.

- Anatole Lefort, Yohan Pipereau, Kwabena Amponsem Boateng, Pierre Sutra, Gaël Thomas. J-NVM: Off-heap Persistent Objects in Java, 28th ACM Symposium on Operating Systems Principles, SOSP 2021, online, 2021.

  Abstract This paper presents J-NVM, a framework to access efficiently Non-Volatile Main Memory (NVMM) in Java. J-NVMoffers a fully-fledged interface to persist plain Java objects using failure-atomic blocks. This interface relies internally on proxy objects that intermediate direct off-heap access to NVMM. The framework also provides a library of highly-optimized persistent data types that resist reboots and power failures. We evaluate J-NVMby implementing a persistent backend for the Infinispan data store. Our experimental results, obtained with a TPC-B like benchmark and YCSB, show that J-NVMis consistently faster than other approaches at accessing NVMM in Java.

- Vitor Enes, Carlos Baquero, Alexey Gotsman, Pierre Sutra. Efficient Replication via Timestamp Stability, EuroSys'21 : Sixteen European Conference on Computer System, Eurosys 2021, online, 2021.

  Abstract Modern web applications replicate their data across the globe and require strong consistency guarantees for their most critical data. These guarantees are usually provided via state-machine replication (SMR). Recent advances in SMR have focused on leaderless protocols, which improve the availability and performance of traditional Paxos-based solutions. We propose TEMPO – a leaderless SMR protocol that, in comparison to prior solutions, achieves superior throughput and offers predictable performance even in contended workloads. To achieve these benefits, TEMPO timestamps each application command and executes it only after the timestamp becomes stable, i.e., all commands with a lower timestamp are known. Both the timestamping and stability detection mechanisms are fully decentralized, thus obviating the need for a leader replica. Our protocol furthermore generalizes to partial replication settings, enabling scalability in highly parallel workloads. We evaluate the protocol in both real and simulated geo-distributed environments and demonstrate that it outperforms state-of-the-art alternatives.

- Daniel Barcelona-Pons, Pierre Sutra, Marc Sánchez-Artigas, Gerard París, Pedro García-López. Stateful Serverless Computing with Crucial, ACM Transactions on Software Engineering and Methodology, Vol. 31, No. 3, Article 39, 2022.

Abstract Serverless computing greatly simplifies the use of cloud resources. In particular, Function-as-a-Service (FaaS) platforms enable programmers to develop applications as individual functions that can run and scale independently. Unfortunately, applications that require fine-grained support for mutable state and synchronization, such as machine learning (ML) and scientific computing, are notoriously hard to build with this new paradigm. In this work, we aim at bridging this gap. We present CRUCIAL, a system to program highly-parallel stateful serverless applications. CRUCIAL retains the simplicity of serverless computing. It is built upon the key insight that FaaS resembles to concurrent programming at the scale of a datacenter. Accordingly, a distributed shared memory layer is the natural answer to the needs for fine-grained state management and synchronization. CRUCIAL allows to port effortlessly a multi-threaded code base to serverless, where it can benefit from the scalability and pay-per-use model of FaaS platforms. We validate CRUCIAL with the help of micro-benchmarks and by considering various stateful applications. Beyond classical parallel tasks (e.g., a Monte Carlo simulation), these applications include representative ML algorithms such as k-means and logistic regression. Our evaluation shows that CRUCIAL obtains superior or comparable performance to Apache Spark at similar cost (18-40% faster). We also use CRUCIAL to port (part of) a state-of-the-art multi-threaded ML library to serverless. The ported application is up to 30% faster than with a dedicated high-end server. Finally, we attest that CRUCIAL can rival in performance with a single-machine, multi-threaded implementation of a complex coordination problem. Overall, CRUCIAL delivers all these benefits with less than 6% of changes in the code bases of the evaluated applications.

- Carlos Segarra, Simon Shillaker, David Goltzsche, Anjo Vahldiek-Oberwagner, Michael Steiner, Mona Vij, Lluis Vilanova, Rüdiger Kapitza, Peter Pietzuch. T-Less: A Confidential Serverless Runtime with Attestation and Authorisation, *Under review*, 2022.

Abstract Serverless cloud applications execute as functions, and their trust-worthiness depends on the underlying serverless runtime. Confidential computing uses hardware-based trusted execution environments (TEEs) to improve the confidentiality and integrity of execution. TEEs, however, require attestation proofs, which assert the integrity of the TEE hardware and the executed code. Current attestation models for TEEs, however, are ill-suited for serverless applications: (i) they assume that attestation is performed ahead of execution, which is infeasible with functions scheduled dynamically by the provider; (ii) they introduce a substantial overhead due to the cost of attestation, which becomes prohibitive with thousands of functions; (iii) they do not support user-defined authorisation poli- cies to control sensitive data disclosure to functions; and (iv) they do not offer users evidence of trustworthy execution.

We describe *T-Less*, a new serverless runtime for confidential serverless applications that combines statically-defined attestation and authorisation policies with dynamic enforcement. When deploying a serverless application, a user provides a *T-Plan* that specifies an application's static function call graph, annotated with attestation and authorisation constraints. T-Less executes each function after successful attestation according to the T-Plan. When function invocations are chained, T-Less collects a sequence of attestation proofs, which must be consistent with the specified call graph. The attestation proofs are also used to authorise functions to access sensitive data based on user-defined policies. After execution, the attestation proofs are aggregated into an evidence graph and returned to the client to demonstrate trustworthy execution.

- Simon Shillaker, Carlos Segarra, Mayeul Fournial, Peter Pietzuch. Granny: Fine-Grained Distribution of Scientific Workloads in the Cloud, *Under review*, 2022.

Abstract Due to their high parallelism and dynamic resource needs, many scientific workloads may benefit from cloud deployments. Today's cloud-based scientific platforms execute jobs on dedicated pools of VMs, resulting in users paying for underutilised resources, and providers being unable to share resources between tenants. While fine-grained resource sharing could address these issues, the programming models used in scientific applications, namely shared

memory and message passing, make this difficult: shared memory applications cannot be distributed across hosts without the significant performance overhead of synchronising data; and message passing employs static sets of stateful processes, which makes it challenging to vary the distribution of long-running jobs.

We propose *Granny*, a runtime that transparently distributes multi-tenant OpenMP/MPI workloads across shared hosts. It supports thread- and process-granular scheduling across a cluster of hosts, using a new software-based isolation abstraction called a *Granule*. Granules interrupt applications at specific *control points*, such as system calls or OpenMP/MPI API calls. At each control point, Granny may allocate new Granules to meet an application's parallelism requirements, or migrate Granules to improve locality. To support distributed shared memory, Granny tracks and synchronises changes to application's shared address spaces; to support message passing, Granny offers asynchronous point-to-point communication between Granules.

- Pedro García López, Aleksander Slominski, Michael Behrendt, Bernard Metzler: Serverless Predictions: 2021-2030 CoRR abs/2104.03075

  Abstract Within the next 10 years, advances on resource disaggregation will enable full transparency for most Cloud applications: to run unmodified single-machine applications over effectively unlimited remote computing resources. In this article, we present five serverless predictions for the next decade that will realize this vision of transparency – equivalent to Tim Wagner's Serverless SuperComputer or AnyScale's Infinite Laptop proposals.

- Germán T. Eizaguirre, Marc Sánchez-Artigas, Pedro García-López: A milestone for FaaS pipelines; object storage-vs VM-driven data exchange Middleware '21: 22nd International Middleware Conference

  Abstract Serverless functions provide high levels of parallelism, short startup times, and "pay-as-you-go" billing. These attributes make them a natural substrate for data analytics workflows. However, the impossibility of direct communication between functions makes the execution of workflows challenging. The current practice to share intermediate data among functions is through remote object storage (e.g., IBM COS). Contrary to conventional wisdom, the performance of object storage is not well understood. For instance, object storage can even be superior to other simpler approaches like the execution of shuffle stages (e.g., GroupBy) inside powerful VMs to avoid all-to-all transfers between functions. Leveraging a genomics pipeline, we show that object storage is a reasonable choice for data passing when the appropriate number of functions is used in shuffling stages.

- Josep Sampe; Pedro Garcia-Lopez; Marc Sanchez-Artigas; Gil Vernik; Pol Roca-Llaberia; Aitor Arjona Toward Multicloud Access Transparency in Serverless Computing IEEE Software, Vol. 38

  Abstract 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.

- Daniel Barcelona-Pons, Pedro García-López Benchmarking parallelism in FaaS platforms Future Generation Computer Systems, Vol. 124

  Abstract Serverless computing has seen a myriad of work exploring its potential. Some systems tackle Function-as-a-Service (FaaS) properties on automatic elasticity and scale to run highly-parallel computing jobs. However, they focus on specific platforms and convey that their ideas can be extrapolated to any FaaS runtime. An important question arises: do all FaaS platforms fit parallel computations? In this paper, we argue that not all of them provide the necessary means to host highly-parallel applications. To validate our hypothesis, we create a comparative framework and categorize the architectures of four cloud FaaS offerings, emphasizing parallel performance. We attest and extend this description with an empirical experiment that consists in plotting in deep detail the evolution of a parallel computing job on each service. The analysis of our results evinces that FaaS is not inherently good for parallel computations and architectural differences across platforms are decisive to categorize their performance. A key insight is the importance of virtualization technologies and the scheduling approach of FaaS platforms. Parallelism improves with lighter virtualization and proactive scheduling due to finer resource allocation and faster elasticity. This causes some platforms like AWS and IBM to perform well for highly-parallel computations, while others such as Azure present difficulties to achieve the required parallelism degree. Consequently, the information in this paper becomes of special interest to help users choose the most adequate infrastructure for their parallel applications.

- Pablo Gimeno Sarroca, Marc Sánchez-Artigas MLLess: Achieving Cost Efficiency in Serverless Machine Learning Training CoRR abs/2206.05786

  Abstract Function-as-a-Service (FaaS) has raised a growing interest in how to "tame" serverless computing to enable domain-specific use cases such as data-intensive applications and machine learning (ML), to name a few. Recently, several systems have been implemented for training ML models. Certainly, these research articles are significant steps in the correct direction. However, they do not completely answer the nagging question of when serverless ML training can be more cost-effective compared to traditional "serverful" computing. To help in this endeavor, we propose MLLess, a FaaS-based ML training prototype built atop IBM Cloud Functions. To boost cost-efficiency, MLLess implements two innovative optimizations tailored to the traits of serverless computing: on one hand, a significance filter, to make indirect communication more effective, and on the other hand, a scale-in auto-tuner, to reduce cost by benefiting from the FaaS sub-second billing model (often per 100ms). Our results certify that MLLess can be 15X faster than serverful ML systems at a lower cost for sparse ML models that exhibit fast convergence such as sparse logistic regression and matrix factorization. Furthermore, our results show that MLLess can easily scale out to increasingly large fleets of serverless workers.

- Josep Sampe, Marc Sánchez-Artigas, Gil Vernik, Ido Yehekzel, Pedro García-López: Outsourcing Data Processing Jobs with Lithops 2018 IEEE/ACM 4th International Workshop on Extreme Scale Programming Models and Middleware (ESPM2)

  Abstract Unexpectedly, the rise of serverless computing has also collaterally started the "democratization" of massive-scale data parallelism. This new trend heralded by PyWren pursues to enable untrained users to execute single-machine code in the cloud at massive scale through platforms like AWS Lambda. Driven by this vision, this article presents Lithops, which carries forward the pioneering work of PyWren to better exploit the innate parallelism of la MapReduce tasks atop several Functions-as-a-Service platforms. Instead of waiting for a cluster to be up and running in the cloud, makes easy the task of spawning hundreds and thousands of cloud functions to execute a large job in a few seconds from start. With Lithops, for instance, users can painlessly perform exploratory data analysis from within a Jupyter notebook, while it is the Lithops's engine which takes care of launching the parallel cloud functions, loading dependencies, automatically partitioning the data, etc. In this article, we describe the design and innovative features of Lithops and evaluate it using several representative applications, including sentiment analysis, Monte Carlo simulations, and hyperparameter tuning. These applications manifest the Lithops ability to scale single-machine code computations to thousands of cores. And very importantly, without the need of booting a cold cluster or keeping a warm cluster for occasional tasks.

- Aitor Arjona, Gerard Finol, Pedro Garcia-Lopez: Transparent Serverless execution of Python multiprocessing applications CoRR abs/2205.08818

  Abstract Access transparency means that both local and remote resources are accessed using identical operations. With transparency, unmodified single-machine applications could run over disaggregated compute, storage, and memory resources. Hiding the complexity of distributed systems through transparency would have great benefits, like scaling-out local-parallel scientific applications over flexible disaggregated resources. This paper presents a performance evaluation where we assess the feasibility of access transparency over state-of-the-art Cloud disaggregated resources for Python multiprocessing applications. We have interfaced the multiprocessing module with an implementation that transparently runs processes on serverless functions and uses an in-memory data store for shared state. To evaluate transparency, we run in the Cloud four unmodified applications: Uber Research's Evolution Strategies, Baselines-AI's Proximal Policy Optimization, Pandaral.lel's dataframe, and ScikitLearn's Hyperparameter tuning. We compare execution time and scalability of the same application running over disaggregated resources using our library, with the single-machine Python libraries in a large VM. Despite the significant overheads of remote communication, we achieve comparable results and we observe that the applications can continue to scale beyond VM limited resources leading to a better speedup and parallelism without changing the underlying code or application architecture.

- Pedro García-López, Marc Sánchez-Artigas, Simon Shillaker, Peter Pietzuch, David Breitgand, Gil Vernik, Pierre Sutra, Tristan Tarrant, Ana Juan-Ferrer, Gerard París: Trade-Offs and Challenges of Serverless Data Analytics Technologies and Applications for Big Data Value

  Abstract Serverless computing has become very popular today since it largely simplifies cloud programming. Developers do no longer need to worry about provisioning or operating servers, and they have to pay only for the compute resources used when their code is run. This new cloud paradigm suits well for many applications, and researchers have already begun investigating the feasibility of serverless computing for data analytics. Unfortunately, today's serverless computing presents important limitations that make it really difficult to support all sorts of analytics workloads. This chapter first starts by analyzing three fundamental trade-offs of today's serverless computing model and their relationship with data analytics. It studies how by relaxing disaggregation, isolation, and simple scheduling, it is possible to increase the overall computing performance, but at the expense of essential aspects of the model such as elasticity, security, or sub-second activations, respectively. The consequence of these trade-offs is that analytics applications may well end up embracing hybrid systems composed of serverless and serverful components, which we call ServerMix in this chapter. We will review the existing related work to show that most applications can be actually categorized as ServerMix.

- Raul Saiz-Laudo, Marc Sanchez-Artigas: EGEON: Software-Defined Data Protection for Object Storage CoRR abs/2206.13162

  Abstract With the growth in popularity of cloud computing, object storage systems (e.g., Amazon S3, OpenStack Swift, Ceph) have gained momentum for their relatively low per-GB costs and high availability. However, as increasingly more sensitive data is being accrued, the need to natively integrate privacy controls into the storage is growing in relevance. Today, due to the poor object storage interface, privacy controls are enforced by data curators with full access to data in the clear. This motivates the need for a new approach to data privacy that can provide strong assurance and control to data owners. To fulfill this need, this paper presents EGEON, a novel software-defined data protection framework for object storage. EGEON enables users to declaratively set privacy policies on how their data can be shared. In the privacy policies, the users can build complex data protection services through the composition of data transformations, which are invoked inline by EGEON upon a read request. As a result, data owners can trivially display multiple views from the same data piece, and modify these views by only updating the policies. And all without restructuring the internals of the underlying object storage system. The EGEON prototype has been built atop OpenStack Swift. Evaluation results shows promise in developing data protection services with little overhead directly into the object store. Further, depending on the amount of data filtered out in the transformed views, end-to-end latency can be low due to the savings in network communication.

3 Open source code repositories

The following is a list of all the open source repositories containing code developed during the project:

- Lithops https://github.com/lithops-cloud/lithops
- Lithops Multi-Cloud Benchmark https://github.com/lithops-cloud/applications/tree/master/benchmarks
- Triggerflow https://github.com/triggerflow/triggerflow
- Crucial https://github.com/crucial-project/crucial
- Crucial DSO https://github.com/crucial-project/dso
- Primula https://github.com/cloudbutton/primula
- MLLess https://github.com/cloudbutton/mlless
- FaaS Parallelism Benchmark https://github.com/cloudbutton/faas-parallelism-benchmark
- Java Hybrid Serverless Executor and Serverless Elastic Algorithms https://github.com/cloudbutton/serverless-elastic-exploration
- Geospatial Workflows https://github.com/cloudbutton/geospatial-usecase
- Lithops Airflow plugin https://github.com/lithops-cloud/airflow-plugin
- JNVM https://github.com/jnvm-project
- Serverless Shell https://github.com/crucial-project/serverless-shell
- Correctness of egalitarian Paxos https://github.com/otrack/on-epaxos-correctness
- Faasm - https://github.com/faasm/faasm
- Faasm OpenMP/MPI support - https://github.com/faasm/faabric
- Faasm Python support - https://github.com/faasm/python
- Faasm MPI experiments - https://github.com/faasm/experiment-mpi
- Faasm OpenMP experiments - https://github.com/faasm/experiment-openmp
- Faasm Confidential Computing support - https://github.com/faasm/experiment-sgx
- Faasm Lithops integration - https://github.com/faasm/lithops
- Infinispan - https://github.com/infinispan/infinispan
- Infinispan Kubernetes Operator - https://github.com/infinispan/infinispan-operator
- Infinispan Quarkus/Native - https://github.com/infinispan/infinispan-quarkus
- Infinispan Cloud Images - https://github.com/infinispan/infinispan-images
- Infinispan Python Client - https://github.com/infinispan/python-client
- SLA management - https://github.com/cloudbutton/sla-management
- Geospatial - https://github.com/matrix-foundation-cb/Geospatial-Cloudbutton

4 Individual exploitation plans

While the project consortium has still been mostly focused on the research and development aspect, the partners have begun exploring the potential avenues for exploitation of the project results in the context of their specific businesses.

4.1 Red Hat

Red Hat is the world's leading provider of enterprise open source solutions, including high-performing Linux, cloud, container, and Kubernetes technologies. Red Hat participates in and creates community-powered upstream projects, by contributing code, leading on standards, collaborating on content, stewarding projects, mentoring leaders, and sponsoring events. The communities drive innovation at a very fast pace, building on the continuous feedback of their users and experimenting with new ideas and implementations. Based on internal assessments, a reduced set of the community driven technologies are integrated, polished and hardened as part of a portfolio of battle tested, mature and multi-year supported products, commercialized together with a rich ecosystem of services and certifications. These products are what ends-up in the hands of Red Hat customers. Red Hat is committed to maintaining a strong, continuous link with the community, creating a healthy, fruitful cycle between community and product.

4.1.1 Business Development possibilities

4.1.1.1 Identification of individual exploitation possibilities for the project

The improvements made to Infinispan in the context of the CloudButton project will have a number of very significant benefits for both the current offerings as well as adoption in new areas. Red Hat currently includes Infinispan in a number of products and services: Red Hat Data Grid (RHDG), Red Hat Enterprise Application Platform (EAP), Red Hat Single Sign On (RHSSO), and more. Several CloudButton-driven technologies and improvements are being integrated into RHDG. This integration has been done at a faster than usual pace due to their usefulness and their alignment to market needs.

4.1.1.2 Short description of key outcomes to be exploited and the innovation potential

Highlights of the work done within the CloudButton efforts:

- the improvements made to the Infinispan Kubernetes Operator, including service provisioning, autoscaling and metrics have made RHDG much more attractive to customers in terms of manageability and monitoring
- the reduced memory footprint and startup time allow for improved performance and higher density deployments, especially in the context of Red Hat's Serverless product, where RHDG will be deployed to offer high-performance in-memory storage alongside other, more traditional persistent storage.
- the new memory scaling algorithm, currently named Anchored Keys upstream, will be particularly useful in the context of Red Hat OpenShift Serverless Eventing and Serving.

4.1.1.3 Potential addressable market & customers

Red Hat Data Grid is part of Red Hat Runtimes, a cornerstone of the Application Services offering. For this reason, it is targeted to all existing and potential Red Hat customers and markets, such as the Financial, Telecommunications, Retail and Public sectors.

4.1.1.4 Timetable for exploitation

The process for features introduced in an upstream project to be delivered as part of a product at Red Hat follows the typical pattern of going through a variable period in which such features are validated by early adopters, called "community bake-time". During such period, features are refined, additional tests are put in place and documentation is written, in order to get the feature ready for product delivery. Red Hat Quality Engineers perform additional tests to verify that the feature does not introduce regressions before the product is released. It is possible to accelerate this process, by labelling the feature as "Technical Preview" in the product, to set realistic customer expectations in terms of readiness. Such features are still fully supported by Red Hat, with the caveat that they might be subject to further evolution and stabilization. Features which are still not deemed ready for product consumption are disabled in RHDG. Red Hat Data Grid is typically released on a twice-yearly cadence. This means that typically a feature will become part of the product within six to twelve months after its introduction in the community.

4.1.1.5 Potential risks, barriers or limitations (if any)

Infinispan is a very mature project, and the changes and improvements that have been made during the CloudButton contributions have been implemented carefully to avoid disruption to the stability of the APIs and behaviour for the users. Additionally, Red Hat has a very strict and well-defined quality assurance process that validates version-to-version improvements, ensuring that no regressions in performance, reliability, resilience occur. Red Hat does not therefore see any potential risks, aside from the usual ones that are inherently present in any complex software product.

4.1.2 Concrete exploitation path

4.1.2.1 Work done and achievements

We can already claim that Infinispan's participation in the project has been a booster for RHDG's usefulness to Red Hat customers. All of the customers running RHDG on OpenShift are taking advantage of the improvements to the operator, which has greatly simplified the management story. The additional improvements to memory management and deployment density has dramatically reduced the resource allocation for some of the most demanding customers who were running hundreds of nodes in production.

4.1.2.2 KPIs to measure the achievement

Red Hat does not have explicit tools in place which could be used to directly measure the achievements derived from the improvements made as part of the CloudButton participation. Since Red Hat Data Grid is no longer sold as an individual product, but as part of the Red Hat Runtimes bundle, it is not possible to measure changes in sales performance. Additionally, from the perspective of both the community and the product, the impact of a particular feature usually manifests itself more as negative feedback in the form of tickets or support requests when issues arise. Such feedback, however, is only negative in its initial denomination, but it is extremely helpful in improving the software. Positive feedback is typically obtained informally, as part of the ongoing supplier-customer relationship.

4.1.2.3 Roadmap

Red Hat will continue its efforts to improve these technologies the communities and, if proven useful, push them to its product line as soon as they are validated by both the community and ad-hoc customer proof-of-concept deployments.

4.2 Atos

Atos SE (Societas Europaea) is a leader in digital services with pro forma annual revenue of circa € 12 billion and circa 100,000 employees in 72 countries. Serving a global client base, the Group provides Consulting & System Integration services, Managed Services & BPO, Cloud operations, Big Data and Cyber-security solutions, as well as transactional services through Worldline, the European leader in the payments and transactional services industry. With its deep technology expertise and industry knowledge, the Group works with clients across different business sectors: Defense, Financial Services, Health, Manufacturing, Media, Utilities, Public Sector, Retail, Telecommunications and Transportation. Atos is focused on business technology that powers progress and helps organisations to create their firm of the future. The Group is the Worldwide Information Technology Partner for the Olympic and Paralympic Games and is listed on the Euronext Paris market. Atos operates under the brands Atos, Atos Consulting, Atos Worldgrid, Bull, Canopy, Unify and Worldline. With the acquisition of Bull, Atos becomes the largest Western European Cloud provider after Amazon. Its cybersecurity and big data portfolios were also significantly enhanced. For more information, visit: `https://atos.net/`.

Atos Research & Innovation (ARI) is the R&D hub for emerging technologies and a key reference for the whole Atos group. With more than 30 years of experience in running Research, Development and Innovation projects, ARI has become a well-known player in the EU context. Our multidisciplinary and multicultural team has the skills to cover all the activities needed to run projects successfully, from scientific leadership to partnership coordination, from development of emerging technologies to the exploitation of project outcomes, with a strong focus on dissemination, innovation adoption and commercialization.

4.2.1 Business Development possibilities

4.2.1.1 Identification of individual exploitation possibilities for the project

Atos foresees a huge potential on SLA commercialization, based on dynamic information and infrastructure monitoring for data intensive applications, as it is a real market need. The SLA manager is currently considered as an asset to be included in Cloud/Edge offerings. Thanks to the work developed within CloudButton, it is considered as a complementarity for the Connected Intelligence offering within the Life Sciences portfolio.

4.2.1.2 Short description of key outcomes to be exploited and the innovation potential

The main asset developed within the project is the SLA Manager, a tool used to ensure the Quality of Service (QoS) of all services deployed across the Continuum. Automatic renegotiation and distributed SLAs have been identified as the main functionalities of interest to be offered to Atos clients.

4.2.1.3 Potential addressable market & customers

As a big company, Atos operates within two markets:

- Internal market: where Atos itself is the main customer of CloudButton results. In this way, SLA Manager will be adopted internally as part of the XaaS (Everything-as-a-Service) portfolio offering, enhancing the current services available.
- External market: in this case, SLA Manager will be included in the commercial portfolio as a single tools or as a complementary one of other products in the portfolio, e.g., pre-installed by default in Bull Sequana devices.

4.2.1.4 Timetable for exploitation

Given the current Market and Technology Readines Level (MTRL), there are two more steps needed to convert the SLA Manager in a commercial solution:

- First step: perform a gap analysis to identify the most convenient time-to-market and develop the appropriate go-to-market strategy.
- Second step: implement the actions depicted in the abovementioned strategy.

4.2.1.5 Potential risks, barriers or limitations (if any)

Barriers found are the common ones for any research result: technological (low TRL) or market (lack of need identified by potential customers) immaturity. However, as it is planned to be included as part of other complementary offerings, it can be considered as an added value of the current portfolio.

4.2.2 Concrete exploitation path

4.2.2.1 Work done and achievements

ARI is a research unit which direcly depends on a horizontal business unit. Thus, the research portfolio is driven by real company needs in order to support the generation of new businesses. For this reason, a thorough internal procedure is followed to:

- Maximise the impact of project results within the company, combining technical results with business analysis, including marketing and business development considerations to support the feasibility analysis of results commercialization.
- Transfer research results to the commercial portfolio, as part of existing solutions or even creating new ones in order to demonstrate the innovation capacity of the company.

During the project lifespan, SLA Manager has been mapped with other Atos solutions, analysing market needs and gaps, and proposing technical integrations. These results have been presented to the business unit where ARI belongs, highlighting the benefits of including such a solution in the current portfolio. As a result of this, an ongoing internal project is starting to draft the architecture of an improved solution to be included as part of the Atos enhanced commercial offering.

4.2.2.2 KPIs to measure the achievement

According to what has being presented before, there are 3 main criteria for considering a successful exploitation of research results:

- Demonstrate the economic viability of the solution.
- Increase MTRL to fulfil the minimum requirement of the company.
- Successfully integrate the SLA Manager with another solution of the company.

4.2.2.3 Roadmap

There are several activities to be performed prior to include CloudButton results within the Atos portfolio:

- Properly map SLA manager with other Atos solutions. Already done.
- Analyse market and demonstrate viability of the integration. Done.
- Propose an enhanced architecture including the new component. Architecture already presented and discussions ongoing.
- Implement updates in the SLA Manager to make it compatible with the previously identified solution. Not yet started.
- Implement the foreseen integration. Not yet started.

Once this final action is performed, the SLA Manager will be automatically included within the commercial portfolio and offered to Atos customers.

4.3 IBM

IBM Cloud Code Engine is the IBM's latest platform for serverless computing. It allows you to submit invocations with large memory footprint that can be executed long amount of time and also expose K8s API. We worked closely with IBM Cloud Code Engine team to integrate and optimize Lithops as IBM Cloud Code Engine interface. As a result, Lithops is officially listed as an API for IBM Cloud Code Engine and listed under official documentation [**?**] of IBM Cloud. In addition we work on various PoC use cases with major customers, who explore Lithops to deploy massive workloads that requires GPU against IBM Gen2 private cloud. We also in process of discussions with Red Hat to include Lithops as a tool to deploy user workloads against Red Hat OpenShit clusters.

4.3.1 Business Development possibilities

4.3.1.1 Identification of individual exploitation possibilities for the project

The capability to provide on demand serverless cloud resources creates a valuable business opportunities for IBM Cloud for customers in that market segment

4.3.1.2 Short description of key outcomes to be exploited and the innovation potential

We worked closely with IBM Cloud Code Engine team to integrate and optimize Lithops as IBM Cloud Code Engine interface. As a result, Lithops is officially listed as an API for IBM Cloud Code Engine and listed under official documentation [**?**] of IBM Cloud.

4.3.1.3 Potential addressable market & customers

Companies and customers that requires heavy analytics on demand with unlimited resources

4.3.1.4 Timetable for exploitation

Lithops already part of official IBM Cloud Code Engine. It's open for exploitation by the customers. With some we have PoC already

4.3.2 Concrete exploitation path

4.3.2.1 Work done and achievements

We worked closely with IBM Cloud Code Engine team to integrate and optimize Lithops as IBM Cloud Code Engine interface. As a result, Lithops is officially listed as an API for IBM Cloud Code Engine and listed under official documentation [**?**] of IBM Cloud. EMBL moved their production workloads to IBM Cloud

4.3.2.2 KPIs to measure the achievement

Real customers exploitation of the capabilities. Many PoC with customers

4.3.2.3 Roadmap

Integrate capabilities in more complex AI frameworks, e.g. involving Ray for data pre-processing

4.4 EMBL

After the successful use in production of Lithops in the Metaspace platform, EMBL has created a startup called SpaceM that is now in incubation process. EMBL's SpaceM startup is now in the incubation phase for 3 years at the incubator BioInnovation Institute (https://bii.dk/), starting in April 2022. The company plans to exploit the METASPACE metabolomics data platform developed the European projects METASPACE, METACELL, and implemented for serverless scalable computing in our project CloudButton.

4.4.1 Business Development possibilities

There are multiple business development opportunities in the space of metabolomics, spatial metabolomics in particular including artificial intelligence. Notably, the big data challenge in this field requires scalable computing infrastructure that is provided by CloudButton. EMBL is currently pursuing the opportunity of creating a platform for spatial single-cell metabolomics for drug discovery. This is enabled by two technologies: spatial metabolomics in the experimental space, and METASPACE in the computing space.

4.4.1.1 Identification of individual exploitation possibilities for the project

Working together with the EMBL transfer department, as well as with the incubator BioInnovation Institute, we have set up the incubation phase of our startup to develop the platform including software. The incubation phase will be supported by 2.4 million Euro funding from the BioStudio program of the BioInnovation Institute specially designed for translating academic developments into business.

4.4.1.2 Short description of key outcomes to be exploited and the innovation potential

The major outcome to be exploited in our startup SpaceM is the scalable serverless implementation of METASPACE (Lithops-METASPACE). This is particularly critical taking into account the ambition of SpaceM to create high-throughput platform that would be generating over 100 datasets per day (compared to 10 datasets currently). Processing these datasets will require scalability and performance that is provided by Lithops-METASPACE.

4.4.1.3 Potential addressable market & customers

The major customers of METASPACE-using startup SpaceM will be large pharmaceutical companies, including Sanofi, Novo Nordisk, AstraZeneca, Merck, BoeringerIngelheim, and Novartis. The market will include preclinical workflows for drug discovery and development, toxicology, target identification, and Drug Metabolism and Pharmacokinetics (DMPK).

4.4.1.4 Timetable for exploitation

The incubation phase is planned for 3 years.

4.4.1.5 Potential risks, barriers or limitations (if any)

A major risk of Lithops-METASPACE related to the commercialization activities in pharmaceutical space is the use of the cloud computing. Historically, pharmaceutical companies were conservative in using cloud, in particular for third-party applications. However, over the past decade this changed and multiple pharmaceutical companies are actively using cloud, and have in place review procedures for third-party cloud software.

4.4.2 Concrete exploitation path

We are currently developing the business plan for our startup SpaceM currently incubated at the BioInnovation institute, including the use of Lithops-METASPACE.

4.4.2.1 Work done and achievements

In the framework of the competitive BioStudio program of the BioInnovation Institute, we have secured 2.4 million Euro for the incubation phase of the startup SpaceM on spatial single-cell metabolomics for drug discovery that will use Lithops-METASPACE as the foundational data analysis and management platform.

4.5 URV

Universitat Rovira i Virgili (URV) is a medium sized University in Tarragona, Spain, URV's CloudLab research group has coordinated the CloudButton project and it has extensive experience in exploitation activities and technology transfer and consulting contracts.

URV is in the process of creating a startup focused on Cloud-based OMICs Data Analytics. We obtained Next Generation Funds in 2022 (150K, DATOMA project) to create a startup and develop the product and business plans. We hired staff that is building the product on top of Lithops. This is a multidisciplinary effort including CloudLab group lead by Pedro Garcia Lopez with Metabolomic group lead by Oscar Yanes.

4.5.1 Business Development possibilities

CloudButton project has created novel business development possibilities in two main fields: (i) migration of legacy applications to the Cloud and (ii) Big Data analytics in the Cloud.

In particular, the implementation of OMICs Big Data pipelines in CloudButton (Metabolomics, Genomics), is opening interesting exploitation opportunities in the research group.

A research group expert in Metabolomics (YanesLab) in Cyber ISCIII (Spain) contacted CLoudLab to move their algorithms to the Cloud in a parallel way for large data volumes. Now both groups are working together joining Cloud expertise (CloudLab-Lithops) and OMICS expertise (YanesLab) to create a new startup. We received Next Generation Funds to build a Proof of Concept and create a startup in 2023.

4.5.1.1 Identification of individual exploitation possibilities for the project

After different consulting and training periods, the main exploitation possibility is to offer a Web-based solution for OMICs data analysts to run complex analytic pipelines in the Cloud with no effort.

The complexity of the Cloud is precluding many institutions and industries to run their code there. We will offer optimized parallel code with Lithops that runs efficiently in the Cloud.

Since our initial achievements outperform significantly commercial solutions in time and cost, our initial target is Metabolomics data analytics. But we plan to extend it to genomic pipelines in the short term.

After receiving some training and consulting sessions, we already have contacts with Big Pharma companies interested in working with us (B2B - Business to Business).

4.5.1.2 Short description of key outcomes to be exploited and the innovation potential

The key outcome to be exploited is: OMICs optimized parallel algorithms moved to the Cloud with Lithops. We will offer them in a Web based platform requiring no coding or Cloud background.

4.5.1.3 Potential addressable market & customers

In principle, we are targeting B2B approaches with a special focus on Pharma industry analyzing and producing OMICs data in the Cloud.

Nevertheless, we also consider B2C approaches in an open web site that may be used by research institutions, Hospitals, and data analysts.

The market is huge since metabolomics alone is moving a big market in Pharma industries, research labs, Hospitals and Universities. And we also plan to incorporate genomic pipelines which will open up the market to more interested institutions and companies.

4.5.1.4 Timetable for exploitation

We received funding to create a MVP during 2022-2023. The product is really advanced and we plan to create the company in 2023.

We plan to start exploitation activities in the end of 2023.

4.5.1.5 Potential risks, barriers or limitations (if any)

A big potential risk is that new companies emerge to compete in the same segment. Cloud OMICs is a hot emerging area with some companies trying to position their products. Big Cloud providers like Amazon are also offering data analytics solutions to their clients.

Nevertheless, existing solutions are complex and require expert staff or very high costs on consulting. Our no-code platform will streamline the execution of specialized computation algorithms in the Cloud.

Another limitation is in the user interface and visualization of the product. We must really provide now a solution that hides all complexity and can reach to many users. If we fail there the user adoption could be slow.

4.5.2 Concrete exploitation path

We are in the early stages of exploitation. We are building a MVP on to of Lithops, and also seeking funding and contacts for an initial B2C demonstration to the community.

The exploitation path starts in the sedond part of 2023, when the product is ready, and we can start business exploitation and commercialization of the solution.

4.5.2.1 Work done and achievements

A MVP is being developed now on top of Lithops. We already moved to the Cloud and optimized a number of algorithms from metabolomics (YanesLab). Algorithms that took hours now are solved in less than 5 minutes in the Cloud with a reasonable cost. We also outperform existing commercial solutions in time and cost. We cannot provide details here for the sake of company's future.

4.5.2.2 KPIs to measure the achievement

The major KPIs to measure the success of this exploitation activity is (i) funds obtained to develop the commercial project and (ii) incomes for the new company in 2023.

We already obtained more than 150K euros from different sources (Next Generation PoC, Xartec, URV). We also continue trying to attract more investment for the company. We will obtain more funds in the next months to develop the idea.

Another clear KPI will be the cash flow of the new company once it is created. We expect to have clients in the Pharma industry very soon and have incomes by the end of 2023.

4.5.2.3 Roadmap

The Roadmap is the following:

- Phase 1 (Product development): We started in January 2022 and an initial MVP should be finished by the end of 2022.
- Phase 2 (Early Adoption and testing): During the first months of 2023 (M1-M6) we will release a public website open to the community to explore the feasibility of the system and to improve the interface. We will start B2C and B2B demonstrations using this portal.
- Phase 3 (Startup creation and commercialization): During the second half of 2023 we will create the company and start commercializing the solution. We will also seek for more invest money like EIC instrument and others to increase the toolkit of available computing algorithms.

4.6 Imperial College London

The exploitation plan of Imperial College London (IMP) is centred around the continued development of the open-source Faasm platform for stateful serverless applications using WebAssembly. IMP will be taking steps to ensure that the public Faasm repository on GitHub (`https://github.`

com/faasm/faasm) remains supported and under active development beyond the end of the project. In addition, IMP is exploring commercialisation opportunities of existing CloudButton technology, in particular Faasm and its extension to support transparently executing high-performance computing (HPC) applications using a serverless model.

4.6.1 Business Development possibilities

4.6.1.1 Identification of individual exploitation possibilities for the project

In the CloudButton project, IMP has developed the Faasm runtime for executing serverless functions compiled to WebAssembly. WebAssembly is a language-independent binary representation that supports a large number of programming languages through the LLVM compiler toolchain. Faasm therefore enables existing applications written in languages such as C, C++, Rust, etc. to be translated to WebAssembly and executed using a serverless model. Since WebAssembly guarantees memory isolation, the Faasm runtime can execute serverless functions with a higher density compared to classical container-based execution models.

Serverless cloud providers are therefore interested in using WebAssembly technology for the execution of serverless functions to improve the efficiency, scalability and performance of their infrastructure. For example, serverless platforms that use WebAssembly technology are uniquely suited to support edge computing use cases in which serverless functions are run on resource-constrained nodes and devices.

WebAssembly is supported through an expanding global community, and it is becoming a high potential technology that is seeing applicability across many cloud-based use cases. As an emerging technology, WebAssembly therefore may impact the design of future serverless platforms in major ways. These trends make Faasm an emerging technology that could be exploited by cloud providers to construct WebAssembly-based serverless execution platforms. The integration of Faasm with Lithops has shown the technical feasibility of using a WebAssembly-based execution back-end for a scalable distributed serverless FaaS platform.

4.6.1.2 Short description of key outcomes to be exploited and the innovation potential

IMP has developed the following software components and plans to explore the potential for further commercial and academic exploitation:

1. **Faasm serverless WebAssembly runtime**: IMP has created the successful Faasm open-source project on GitHub. Faasm is a high-performance stateful serverless runtime. It provides multi-tenant isolation, yet allows functions to share regions of memory. These shared memory regions support low-latency concurrent access to data, and are synchronised globally to offer large-scale parallelism across multiple hosts.

   Faasm combines software fault isolation from WebAssembly with standard Linux tooling to give security and resource isolation at low cost. It can run serverless functions side-by-side as threads of a single runtime process, with lower overheads and faster boot times compared to traditional Linux containers.

   Faasm defines a custom host interface that extends the WASI standard to include function inputs and outputs, chaining functions, managing state, accessing the distributed filesystem, dynamic linking, and pthreads.

2. **OpenMP/MPI programming support for Faasm serverless applications**: Faabric is a library on top of Faasm that provides scheduling, messaging and state for distributed serverless runtimes. It offers an implementation of distributed programmning models that can be mapped to underlying stateful serverless threads.

   Currently, Faabric supports a shared memory programming model, as implemented by the OpenMP API. Existing OpenMP applications can be executed using Faabric in a serverless fashion. In addition, it also support distributed message passing, according to the MPI model. It provides a suitable implementation of a communication layer between stateful serverless functions to emulate message passing semantics.

4.6.1.3 Potential addressable market & customers

The Faasm technology can play a role in future resource-efficient serverless platforms, as rolled out by large public cloud providers. Faasm is based on the open WebAssembly standard, and it is likely that serverless computing models based on WebAssembly will increase their market share in the future. They offer performance, interoperability, language independence and scalability.

4.6.1.4 Timetable for exploitation

IMP has promoted the Faasm technology developed in CloudButton and, more generally, WebAssembly to major public cloud providers. IMP is in discussions regarding potential exploitation paths, and how to grow the Faasm ecosystem. There is at least one commercial effort in which a large technology company has implemented parts of the Faasm design as part of their production serverless setting. Further such adoption should create pressure in the marketplace for more cloud providers to provide a WebAssembly-based serverless offering.

4.6.1.5 Potential risks, barriers or limitations (if any)

WebAssembly is a new technology and therefore lacks the maturity of e.g. container-based solutions. There is therefore a risk that the standardisation progress of WebAssembly remains slow, hampering efforts to build serverless platform according to a WebAssembly model. Given the recent resources that went into WebAssembly standardisation work, this risk can be considered to be low.

4.6.2 Concrete exploitation path

4.6.2.1 Work done and achievements

IMP has two exploitation pathways that it is pursuing:

1. The Faasm open-source project on GitHub has attracted substantial interest and independent third-party development, as can be seen by its over 680 stars and 50 forks on GitHub. Faasm enjoys wide visibility in the WebAssembly serverless community, e.g. being mentioned at the Cloud Native 2022 WASM day. The original developer behind Faasm, Simon Shillaker, has joined a European cloud provider to lead their serverless effort and will continue to manage the Faasm open-source community.

2. IMP is exploring further commercial exploitation of the Faasm technology. It is currently negotiating follow-on funding with industrial partners that will develop new use cases on top of Faasm. For example, with Intel, IMP has added support for Confidential Computing using Intel SGX to Faasm. In addition, IMP is in discussions with a North American VC Fund (Amplify Partners) about how the Faasm and WebAssembly serverless ecosystems can be supported further.

4.6.2.2 KPIs to measure the achievement

We will continue to monitor carefully the growth of the Faasm community on GitHub. If the growth continues at the same rate, with further industrial and academic contributions to the project, WebAssembly-based serverless platforms become a viable alternative that will be considered by more cloud providers.

4.6.2.3 Roadmap

We will continue our discussion with industrial partners, potential investors and community members after the end of the CloudButton project. In particular, Simon Shillaker will spend resources to maintain the Faasm project and grow its reach by maintaining the community.