# READDAPT – security and Energy Aware DAta management in cloud/edge/fog/serveless comPuTing


## Abstract

The READAPT secuRity and Energy Aware DAta management in cloud/edge/fog/serveless comPuTing aims to propose architectures and algorithms to provide multi-objective data optimization, considering conflicting objectives such as the CO2 footprint or the privacy levels. In particular, this PhD thesis will focus on data management relying on Serverless Computing so as to save energy consumption while not compromising the security level. On one hand it will consider ephemeral data management to accelerate query processing. On the other hand, it will take into account various storage strategies from clear data on files to blockchain-based solutions or using encryption.


## Intérêt de la collaboration

L'IRISA apporterait son expertise autour de l'optimisation de la gestion de données massives, la supervision de sécurité (des réseaux) et l'utilisation de l'informatique en nuage et sans serveur. Le LIRIS a une expertise complémentaire sur le partage de données sécurisée et l'évaluation de performances des systèmes NoSQL. Enfin, l'Université d'Arizona est reconnue pour ses travaux sur les bancs d'essais dédiés sécurité. L'IRISA et le LIRIS d'une part, le LIRIS et l'Université d'autre part collaborent depuis de nombreuses années. Cette thèse permettrait d'étendre ces collaborations et de permettre des échanges directs entre l'IRISA et l'Université d'Arizona et notamment de permettre des mobilités aux Etats-Unis pour d'autres étudiants de l'Université de Rennes.

## Subject

### Context

Next-generation mobile networks (6G and beyond) generate large amounts of data that are exploited through Big Data platforms for performance or even security purposes. Indeed, the state of the art shows the use of these platforms (e.g. Thales Punch Platform [1]) to collect and analyze data from several heterogeneous sources in order to monitor, for example, network security.

Unfortunately, Big Data also raise some drawbacks. On one hand Big Data has a negative environmental impact. Indeed, data centers that are used to store and process data use a lot of energy. This contributes to climate change. On the other hand, Big data may contain sensitive information, like personal, medical or financial data. these large amounts of data can also lead to a loss of privacy, since they can be used to track movements and/or activities of individuals. In addition, in order to improve response time and scalability, these platforms are evolving towards a decentralized approach by including, in addition to central Cloud servers, the capabilities of other nodes (Edge,Mobile Edge, Fog)[2]. This implies security issues while manipulating data (processing, storage and exchange) to protect Big Data platforms against possible attacks (intrusion, node compromise, etc.).

## Problem

To enhance performances of Big Data platforms, Serverless Computing [3] has been recently proposed as a parallel and distributed environment with high elasticity and thus better scalability and response time as well as reduced bandwidth. The main objective of this PhD thesis is to provide a new secured and energy aware Big Data platform including Edge and Fog capabilities.

The target platform could be inspired by the Thales open-source Punch platform [1] and have a hybrid deployment using the available Edge / Fog equipment and devices and even Cloud ones if needed.

We will also provide the needed security level for that Big Data analytics system in its main steps: storage, treatments and communications. As for the communication and storage security, data must be protected efficiently (encryption, authentication, etc.) according to the context's needs [7]. Securing data storage and exchange using a blockchain-based approach seems to be a relevant alternative to traditional protection like those used with Apache Spark (encryption for RPC connection, local storage encryption, etc.) [8]. Moreover, this type of approach is perfectly suited for distributed systems which is the case of the targeted platform [9].

In this work, we can consider a concrete use case like security supervision. The experimentation of the implemented prototype will be carried out in order to evaluate some performance parameters like the response time (including processing time and transmission delay) and the network resources consumption (bandwidth).

## Organization

The following steps will be considered:

State of the art: the candidate will have to cover the literature on (big) data management, data sharing, distributed system (serverless computing), performance analysis and security monitoring.

Proposal: an architecture with the associated algorithms and data structures will be designed so as to provide the appropriate trade-off between energy consumption and security. Ephemeral data storage will be of particular interest since it can help to accelerate query processing but can also lead to some data linkages. Some questions to be address will be: what data to store? How to store these data (data model, hardware, with or without encryption)?

Validation: in this work, we will consider a security monitoring application. The experimentation of the proposed solution will be carried out in order to evaluate the following parameters: response time (including processing time and transmission delay), network resources consumption (bandwidth), overall energy consumption (Fog, Edge and Cloud equipment), scalability, robustness of the provided security. This evaluation will allow us to compare our solution to existing ones (in particular Spark on HDFS or MinIO) in terms of the above-mentioned criteria. The obtained results will be qualified on the basis of the compromise made between security level, performance level and impact on energy consumption. Due to the lack of real data and queries on the domain, researches will be conducted to provide a security monitoring benchmark.

## Bibliography

[1] The Punch: Cloud native, Simple, Serverless, Thales PunchPlatform. https://punchplatform.com

[2] R. Venanzi, S. Dahdal, M. Solimando, L. Campioni, A. Cavalucci, M. Govoni, M. Tortonesi, L. Foschini, L. Attana, M. Tellarini, C. Stefanelli, « Enabling adaptive analytics at the edge with the

Bi-Rex Big Data platform », Computers in Industry, Volume 147, 2023, https://doi.org/10.1016/j.compind.2023.103876.
[3] Michail Georgoulakis, Laurent d'Orazio, Verena Kantere: From Cloud to Serverless: MOO in the new Cloud epoch. EDBT 2022: 1-4
[4] T. Sylla, M.A. Chalouf, F. Krief, K. Samaké, « Context-Aware Security in the Internet of Things: A
survey », International Journal of Autonomous and Adaptive Communications Systems, IJAACS, 2021. doi: https://doi.org/10.1504/IJAACS.2021.117808
[5] Apache Spark Security, « Unified engine for large scale data analytics », https://spark.apache.org/docs/latest/security.html
[6] A. Razzaq, S.A.H. Mohsan, S.A.K. Ghayyur, M.H. Alsharif, H.K. Alkahtani, F.K. Karim, S.M. Mostafa,
« Blockchain-Enabled Decentralized Secure Big Data of Remote Sensing », Electronics. 2022; 11(19):3164. doi: https://doi.org/10.3390/electronics11193164