

Leading  
the Big Data  
Revolution

### Jean-Thomas Acquaviva

DDN (coordinator)

“

### E PUR SI MUOVE!

**As the convergence of HPC and Big Data has appeared an elusive if not impossible promise, both technologies remaining distant, frozen and immobile in their respective markets: the Evolve project in a very pragmatic way is actually pushing things and... Yes, it moves!**

Through the injection of HPC technologies such as hardware counter monitoring and NUMA allocation, we improved the workload execution of containers for node consolidation, with an average speed-up of 20%. Without altering the simplicity of deploying containers, by paying attention to the ‘bare-metal’ aspects, Evolve provides acceleration. Still in this idea of getting more performance out of the platform, Evolve is investigating and implementing scheme for GPU sharing, where a single GPU can be shared between jobs. This fine grain resource allocation scheme is an important step forward in terms of data center efficiency, for CAPEX as well as OPEX.

The use cases are taking advantage of the platform, not only from the hardware standpoint but of course from the burgeoning converged software stack. Our partners in the automotive market are able to train neural networks over one hundred million data base records. More importantly use case providers can now deploy their containerized big data pipeline on the HPC testbed with the ability to reach extreme level of performance.

Data management is making progress as well with the implementation of Spark over flash native storage now in production.

At last, and we are really proud of it, the complete Data Life Cycle Framework from IBM supported and partially developed within the Evolve project is now Open Source.

Bringing back value to the community is a key focus of the Evolve project... **Let's keep moving!**

”

Read our newsletter and get to know more about our **use cases** and **technology**:



## USE CASES

### A Consolidated View on the Technological Profile and Progress Update for all Use Cases

[Read article](#)

Seven pilots are challenging EVOLVE large-scale testbed with varying needs and goals. The applications are supporting actual business cases into diverse domain areas such as agriculture, preventive machine maintenance, public transport, maritime surveillance, mobility services, image change detection, automotive data-driven systems. But which are the challenges' concerns?

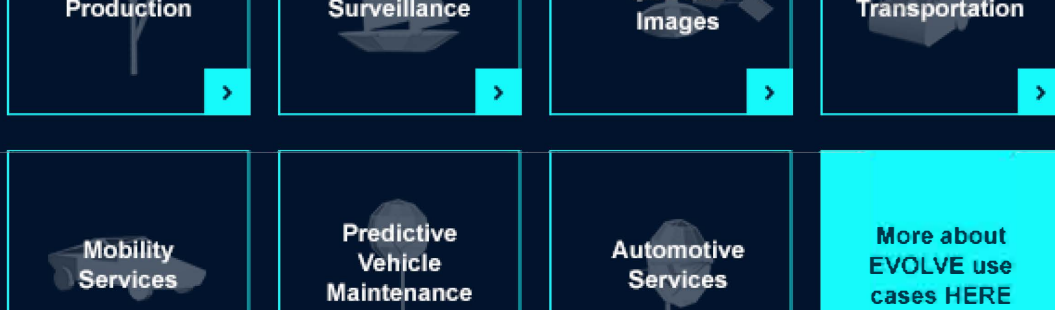
Author - **Vassilis Spitadakis (Neurocom Luxembourg)**

### A Predicting Vehicle Maintenance using on-board and off-board data

[Read article](#)

In this article, we are going to explain how we use data science to develop connected car applications, focusing on predictive maintenance based on DTC codes and available OBD car data.

Author - **Alen Kopic (Koola.io)**



## TECHNOLOGY

### Karvdash: A Kubernetes dashboard for data-science workflows

[Read article](#)

To facilitate high-level user interaction with the Kubernetes execution environment used in EVOLVE, we present Karvdash: a service management software for Kubernetes, which runs in Kubernetes as a service itself.

Author - **Antony Chazapis and Angelos Bilas (FORTH-ICS)**

### Dataset Lifecycle Framework: the swiss army knife for data source management in Kubernetes

[Read article](#)

Hybrid Cloud is rapidly becoming the go-to IT strategy for organizations seeking the perfect mix of scalability, performance and security. As a result, it is now common for an organization to rely on a mix of on-premise and cloud solutions, or “data-sources”, from different providers to store and manage their data...

Author - **Christian Pinto and Yiannis Gkoufas (IBM Research Europe (Ireland Lab))**

### Exploration of GPU sharing policies under GEMM workloads

[Read article](#)

Lately, cloud computing has seen explosive growth, due to the flexibility and scalability it offers. The ever-increasing computational demands, especially from the machine learning domain, have forced cloud operators to enhance their infrastructure with acceleration devices, such as General-Purpose (GP)GPUs or FPGAs...

Author - **Ioannis Oroutzoglou, Dimosthenis Masouros, Konstantina Koliogeorgi, Sotirios Xydis, Dimitrios Soudris (ICCS, NTUA)**

### Interference-aware Orchestration in Kubernetes

[Read article](#)

Nowadays, there is an increasing number of workloads executed on the Cloud. Although multi-tenancy has gained a lot of attention to optimize resource efficiency, current state-of-the-art resource orchestrators rely on typical metrics, such as CPU or memory utilization, for placing incoming workloads on the available pool of resources, thus, neglecting the interference effects from workload co-location. Within EVOLVE, we have designed an interference-aware cloud orchestrator, based on micro-architectural event monitoring. We integrate our solution with Kubernetes achieving higher performance, up to 32% compared to its default scheduler, for a variety of cloud workloads.

Author - **Achilleas Tzenetopoulos, Dimosthenis Masouros, Sotirios Xydis, Dimitrios Soudris (ICCS, NTUA)**

## More about EVOLVE testbed

[Visit page](#) →

Follow the project

[www.evolve-h2020.eu](http://www.evolve-h2020.eu)

Visit us on Social Media



The Consortium



This project has received funding from the European Union Horizon 2020 Research & Innovation Programme under Grant Agreement no. 825061