

e volve

EVOLVE

**Heterogeneity and convergence for Cloud, Big Data
and HPC**

Jean-Thomas Acquaviva
DDN Storage
jt.acquaviva@gmail.com
www.evolve-h2020.eu

Nov. 2021

Project Overview: Heterogeneity and convergence for Cloud, Big Data and HPC

ICT11 - H2020

19 Partners

€M14 Budget

3 years, termination Dec. 2021

Innovation Action: Focus on Innovation
and Industry



DDN Storage
www.ddn.com



BULL
www.atos.net



IBM
www.ibm.com



FORTH
www.ics.forth.gr



Sunlight
www.sunlight.io



Institute of Communications
and Computer Systems
www.microlab.ntua.gr



MemoScale
www.memoscale.com



webLyzard technology
www.webyzard.com



LOBA
www.loba.pt



Thales Alenia Space
www.thalesgroup.com



Space Hellas
www.space.gr



CybeleTech
www.cybeletech.com



Neurocom Luxembourg
www.neurocom.eu



MemEX
www.memexitaly.it



Tiemme SPA
www.tiemmespa.it



Virtual Vehicle
www.v2c2.at



AVL List GmbH
www.avl.com

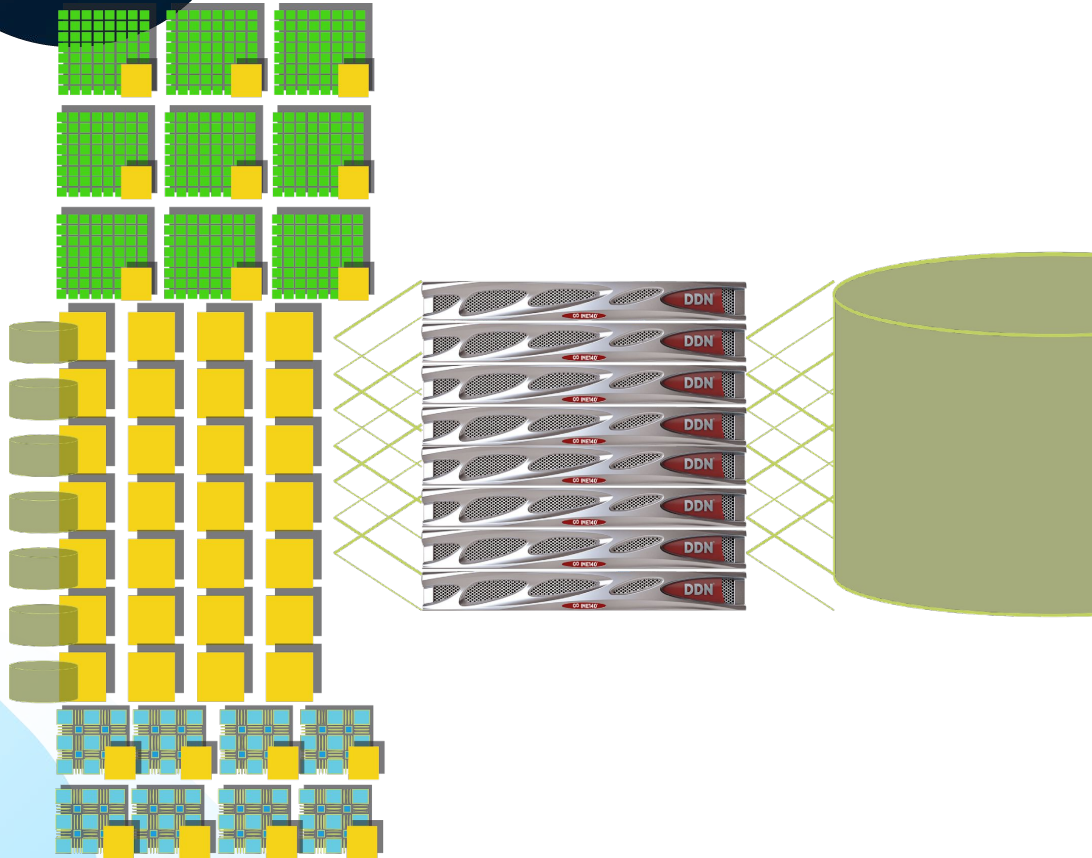


BMW AG
www.bmw.com

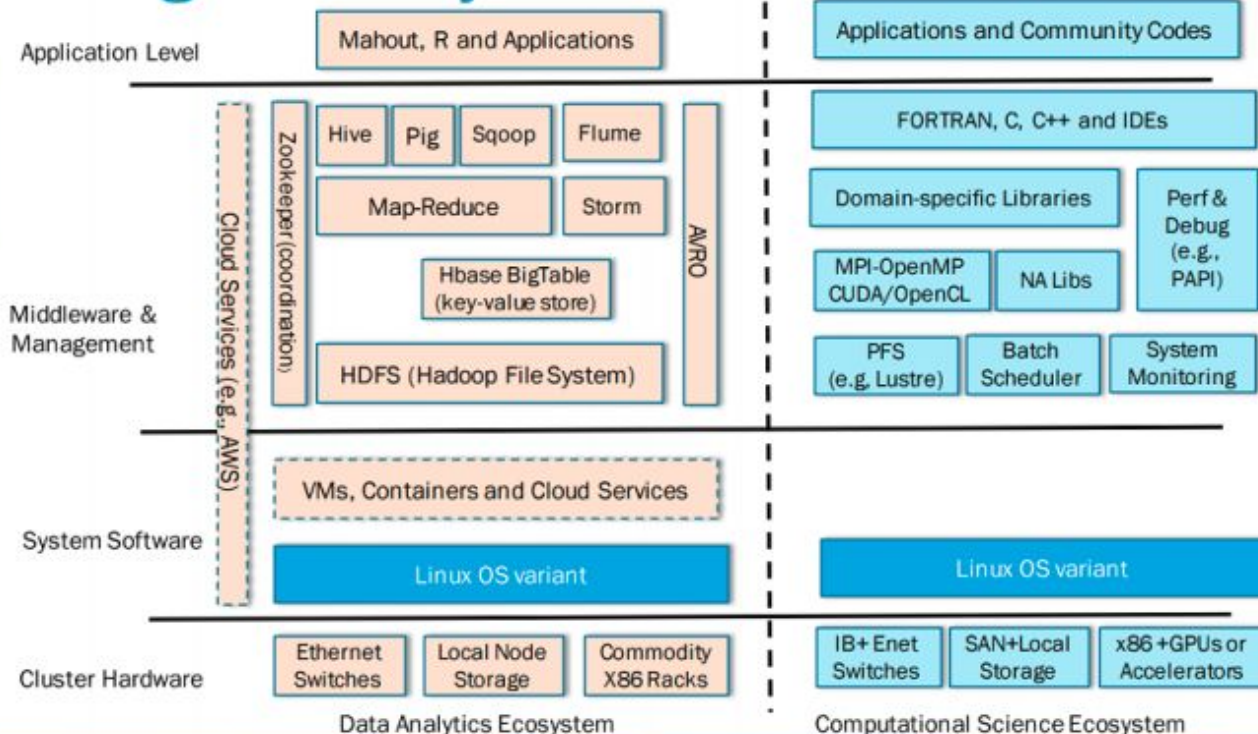


KOOLA
www.koola.io

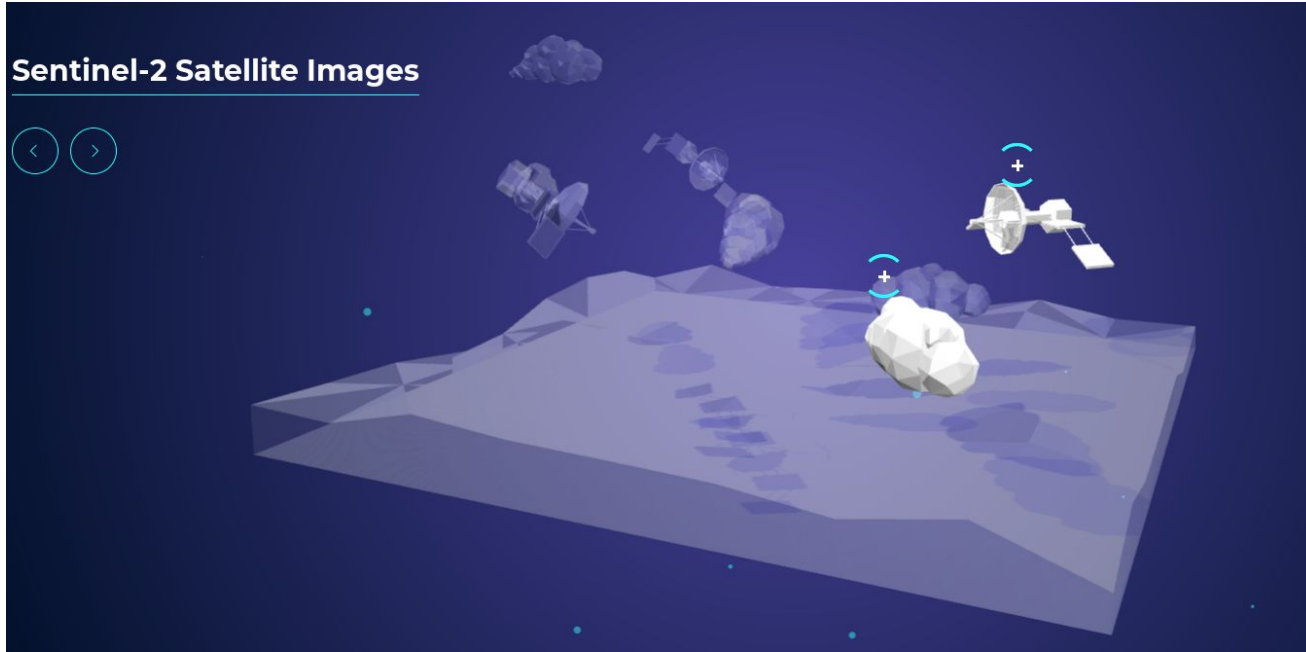
Heterogeneous Platform



Divergent ecosystems



Evolve Use Cases: Satellite



CybeleTech

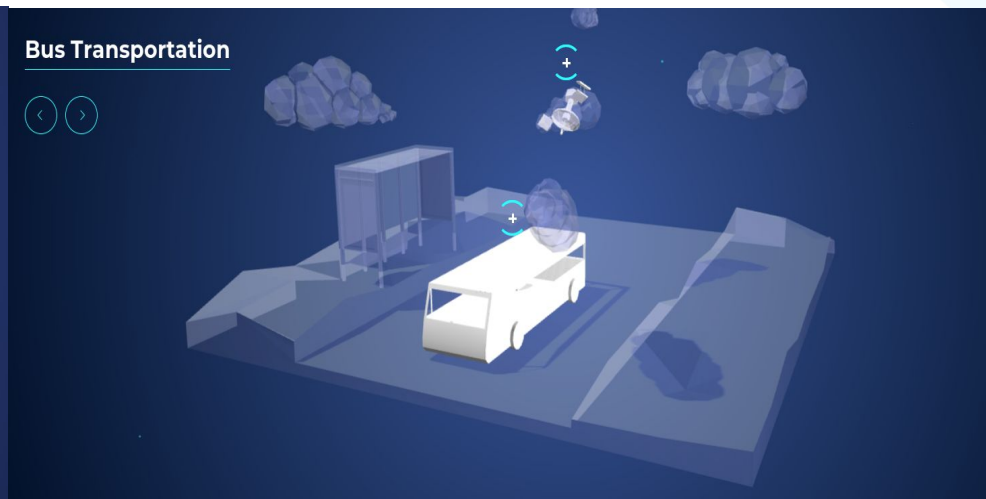
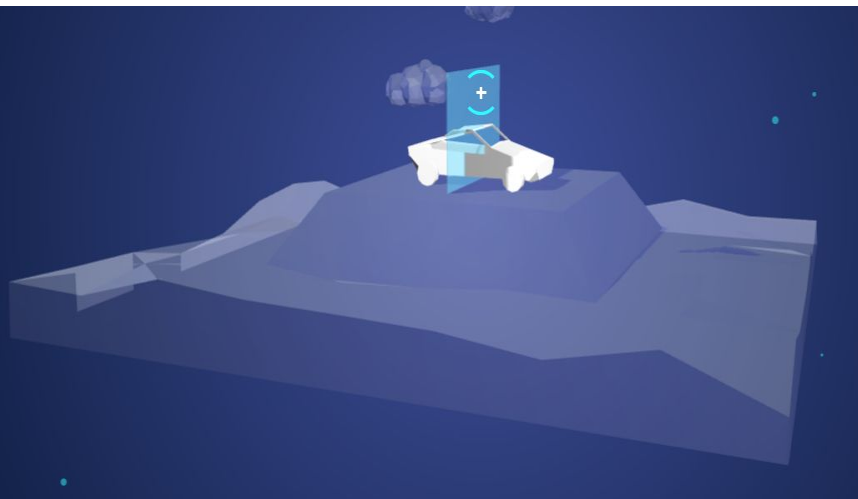
Agriculture optimization

Thales AleniaSpace

Change Detection
at European Scale

SpaceHellas

Maritime Surveillance



AVL Vehicle Predictive Maintenance

Koola AI for automotive Service

BMW Mobility Service

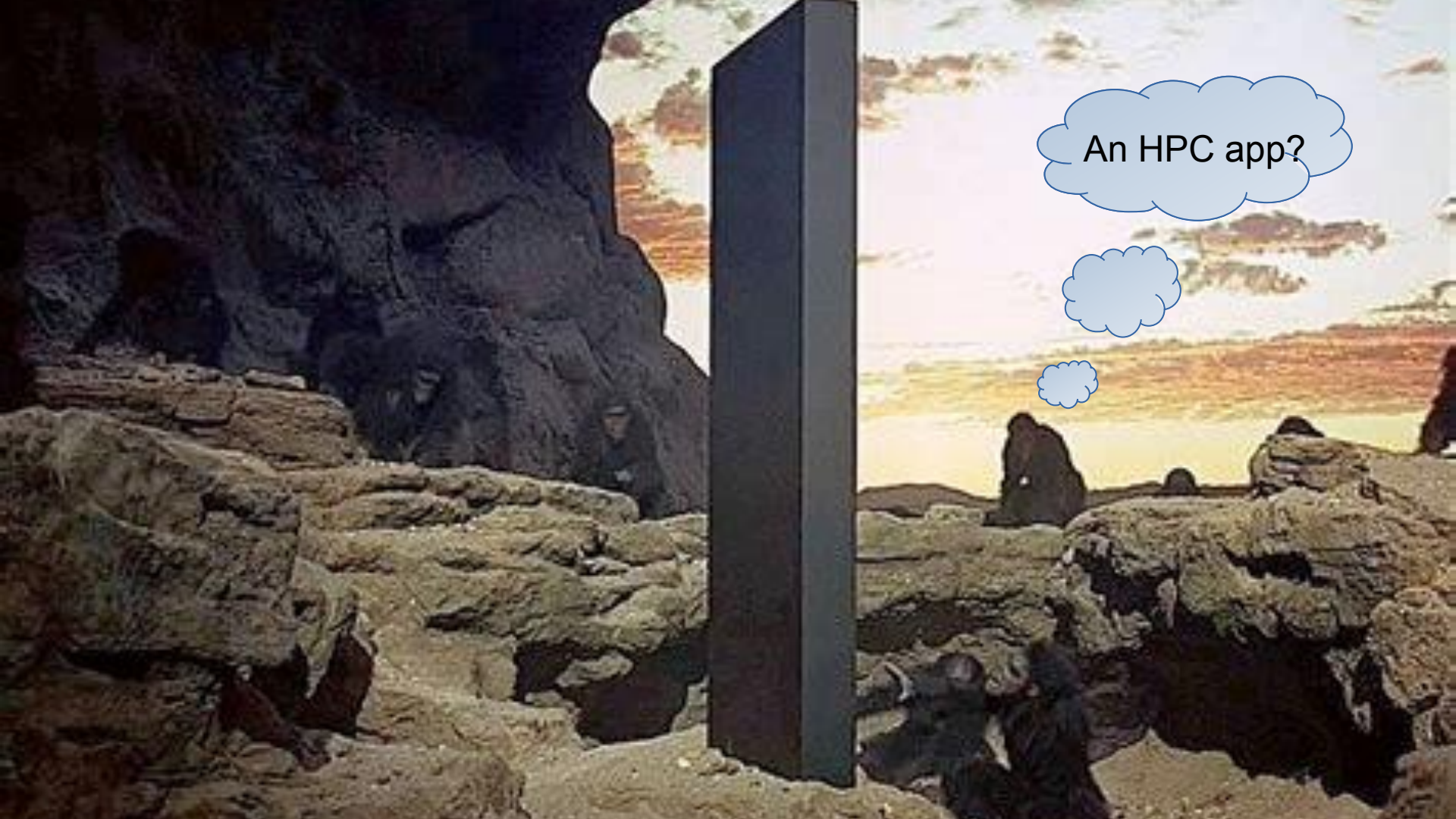
Virtual Vehicle Driver Observation

Tiemme / MeMex Monitored and Improved Service Planning

Concept of the Proof-of-Concept (PoCs)

- Use cases are almost co-designed: engineering work both on Application and the platform
- PoCs are application external from the project, either from partners or external organizations
- PoCs are customers of the platform
 - Test and Try
 - Qualify the User Experience
 - Assess the Porting Effort





An HPC app?

Application heterogeneity

Custom Evolve Interpreter

- Simplified generation of multi-stage workflows for Argo engine
- Support for control state of the K8s cluster
- Manage container images in docker repository
- Manage data files via Python or shell interpreter



The screenshot shows a Zeppelin Notebook interface. The browser address bar indicates the URL is `192.168.30.40:8081/#/notebook/2E0D4XZFS`. The page title is `/Evolve Zeppelin Tutorial/E...`. The `Settings` section is visible, showing the `Interpreter binding` configuration. Below this, there is a list of interpreters with their respective languages and configurations:

- `python %python (default) %python %sq %conda %docker`
- `spark %spark %spark-sq %spark-dip %spark-pyspark %spark %spark-r`
- `md %md`
- `sh %sh`
- `evolve %evolve %evolve-python %evolve-sq %evolve-conda %evolve-docker`
- `file %file`
- `angular %angular`

EVOLVE Result: Accelerators accelerate!

Speed-up

- Some application get x100 speed-up













Performance boost to big data components

- >3x due to better storage + networking capabilities to data IO critical stages
- >10x due to the availability of parallel multi-core processing to compute-heavy tasks

Productivity boost to HPC+Big Data hybrid pipelines

- 10x-30x fewer lines of code through the use of the Big Data ecosystem (Spark, Tensorflow etc)
- More than 10x fewer lines of code due to simpler workflow submission process (notebooks, argo workflows, Kubernetes as opposed to SLURM scripts, manual SSH-based setups etc)

EVOLVE: 12 innovations produced

| | | | |
|---|--|---|---|
|  <p>Play Video</p> |  <p>Argus Holistic performance monitoring system for heterogeneous platform</p> <p>Play Video</p> |  <p>vizcon Visualization and Centralization Micro-Service</p> <p>Play Video</p> |  <p>Cybel Integrating Remote Sensing to Maritime Surveillance</p> <p>Play Video</p> |
|  <p>accelx Hardware and software architecture to support a scalable and interoperable integration of computing acceleration technologies</p> <p>Play Video</p> |  <p>Earth Observer</p> <p>Play Video</p> |  <p>Celso Mobility Dashboard Real-time Public Transportation Monitoring</p> <p>Play Video</p> |  <p>Skynet an adaptive resource allocator for datacentre workloads</p> <p>Play Video</p> |
|  <p>MemoPort</p> <p>Play Video</p> |  <p>ikarvdash High-productivity and performance software stack for data science on Kubernetes</p> <p>Play Video</p> |  <p>ParSiDeTec</p> <p>Play Video</p> |  <p>acritas Maritime Surveillance Platform powered by Oracle</p> <p>Play Video</p> |

Innovations for real!



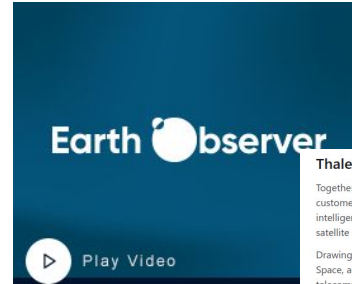
Open source project
datashim.io hosted by the
Linux Foundation



Reduced S3 traffic by
60% on Satellite
Images



Check Climate.gov
climate awareness
website from US Gov.



Check Azure Market Place

Thales Alenia Space partnership

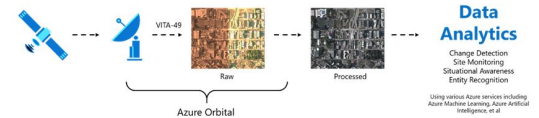
Together with Thales Alenia Space, we are bringing near real-time geospatial data processing capabilities to our customers. The power of processing data on the cloud with Azure Orbital coupled with the application of artificial intelligence (AI) and machine learning technology helps our customers analyze environmental changes captured with satellite imagery.

Drawing on over 40 years of experience and a unique combination of skills, expertise, and cultures, Thales Alenia Space, a Joint Venture between Thales (67 percent) and Leonardo (33 percent), delivers cost-effective solutions for telecommunications, navigation, Earth observation, environmental management, exploration, science, and orbital infrastructures. Governments, institutions, space agencies, and telecom operators count on Thales Alenia Space to design satellite-based systems that provide anytime, anywhere connections and positioning, monitor our planet, enhance management of its resources, and explore our Solar System and beyond.

"Thales Alenia Space and Microsoft are innovating together by combining their expertise in space and cloud technologies. Customers can now combine all the functionality of Thales Alenia Space's DeeperVision solution for processing dataflows and generating timely information with the cloud capabilities of Azure Orbital. This information is enriched by high-speed, high-volume artificial intelligence and machine learning to create an unprecedented impact on and beyond the planet!" - Clarence Duflocq, Vice President Strategy & Innovation, Thales Alenia Space

Use case: Geospatial data value-chain

Customers can use Azure Orbital and Orbital ground stations to bring geospatial data from their satellites. The raw satellite data can then be processed at-scale on the cloud for analysis using various Azure services to achieve goals like Change Detection, Site Monitoring, Situational Awareness, and Entity Recognition.



EVOLVE Scientific Output

- **36 conference / Journal papers**
 - including Usenix and IEEE/ACM journals
- **Book chapters**
- **2 White papers**
- **3 Cookbooks**

EVOLVE Community Contribution

- **+100,000 LoC on GitHub**

Harnessing Heterogeneous Hardware

- GPU, CPU, FPGA, multi-tiered Storage

Software Stack to conceal / Leverage Heterogeneity

- How to improve orchestration between two components, how to debug a workflow.

Demonstrating and Field Experience

- Pilot Application: Co-Design
- Proof-of-Concept: External applications injected into the platform

Consortium



DDN STORAGE
www.ddn.com



BULL
www.atos.net



IBM
www.ibm.com



FORTH
www.ics.forth.gr



OnApp
www.onapp.com



Institute of communications and computer systems
www.microlab.ntua.gr



MemoScale
www.memoscale.com



webLyzard technology
www.weblyzard.com



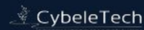
LOBA
www.loba.pt



Thales Alenia Space
www.thalesgroup.com



Space Hellas
www.space.gr



CybeleTech
www.cybeletech.com



Neurocom Luxembourg
www.neurocom.eu



MemEX
www.memexitaly.it



Tiemme SPA
www.tiemmespa.it



Virtual Vehicle
www.v2c2.at



AVL List GmbH
www.avl.com



BMW AG
www.bmw.com



KOOLA
www.koola.io

www.evolve-h2020.eu
info@evolve-h2020.eu



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825061