

# Ryax Data Engineering Platform

## Usage and Internals

**Yiannis Georgiou**  
Co-Founder & CTO



# Ryax Technologies Company

Ryax Technologies was founded in November 2017 with one vision:



***Data must be treated at the right place at the right time***

- We provide a **data engineering software platform** that enables the **data analytics workflows task automation** upon hybrid **edge-cloud** infrastructures
- Background in HPC systems and in particular resource management & scheduling
- Co-Founders used to work at Bull/Atos for 10+ years responsible for SLURM related R&D

# We talk about DataScience!

Data science is a broad field that refers to the collective processes, theories, concepts, tools and technologies that enable the review, analysis and extraction of valuable knowledge and information from raw data

***Technopedia***



**Data  
Science is a journey !!**

# Data Science is a necessary evil

**79**

% of enterprise execs say that companies who don't embrace BigData will lose market strength & may face extinction

*Accenture*

**15**

% Of Data Science projects deploy to production

*Gartner*

**1**

Major reason: Put data pipeline in production is too complicated

*Gartner*

**4**

Data engineers per Data scientist

*CTOvision*



# Data Engineer: the key competency

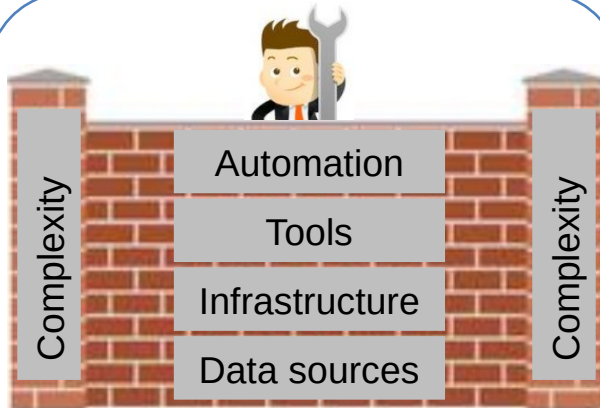
## And extremely rare !!

5X more Data Engineer opening jobs than Data Scientist ones ([Glassdoor](#))



### A CEO

Acquire Market Share  
Address new market  
Be more profitable



### A DataScientist

Data analysis  
Model definition  
Based on standard  
libraries ( Tensorflow,  
DeepLearning, ETL,  
vizualisation ..)



### A Data Engineer



# Our mission

***With Ryax Technologies, finally make companies Data Science profitable***

We provide a data engineering platform that will allow:



To put  
companies  
Data Science  
in production  
in few days



To take into  
account  
companies  
specificities

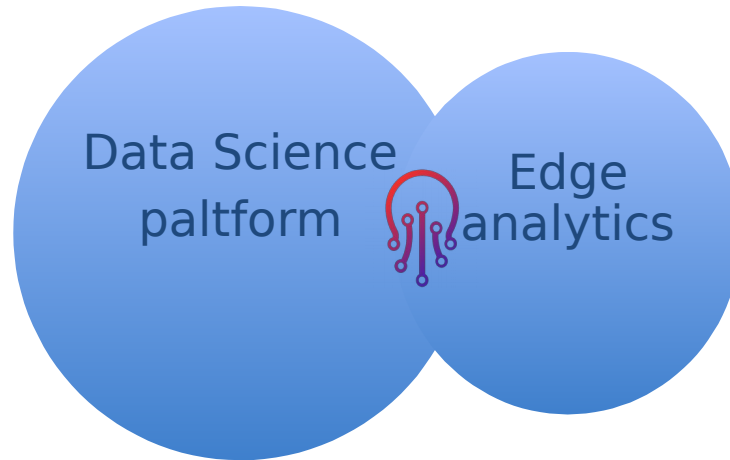


To be in the  
starting  
blocks for the  
next Big Data  
steps

# Our focus

Data engineering tasks of **Near Real Time services**  
(Connected vehicles, digital maintenance ...):

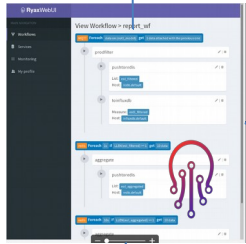
- Hot data (sensors, video & audio streams)
- Massive treatment at the Edge



# Ryax: Data Engineering platform



1. Build your workflow using your own code or provided tools
2. Explain your constraints (security, performances, location...)
3. Run it



Data Science tools (Deep Learning, Machine Learning, IA ...) & APIs (Storage, web services, open data, IOT ...)



Customer Data Science application



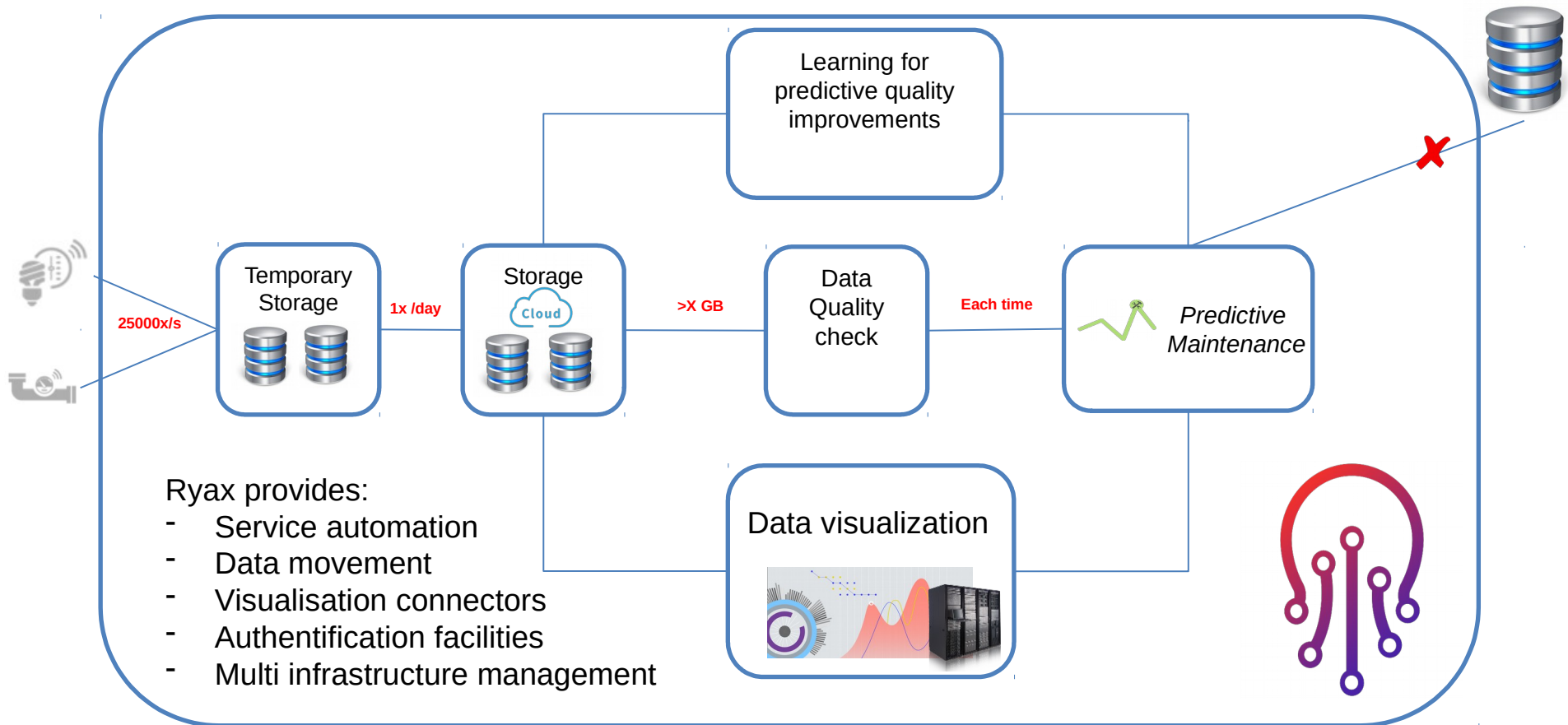
Constraints description using comprehensive language



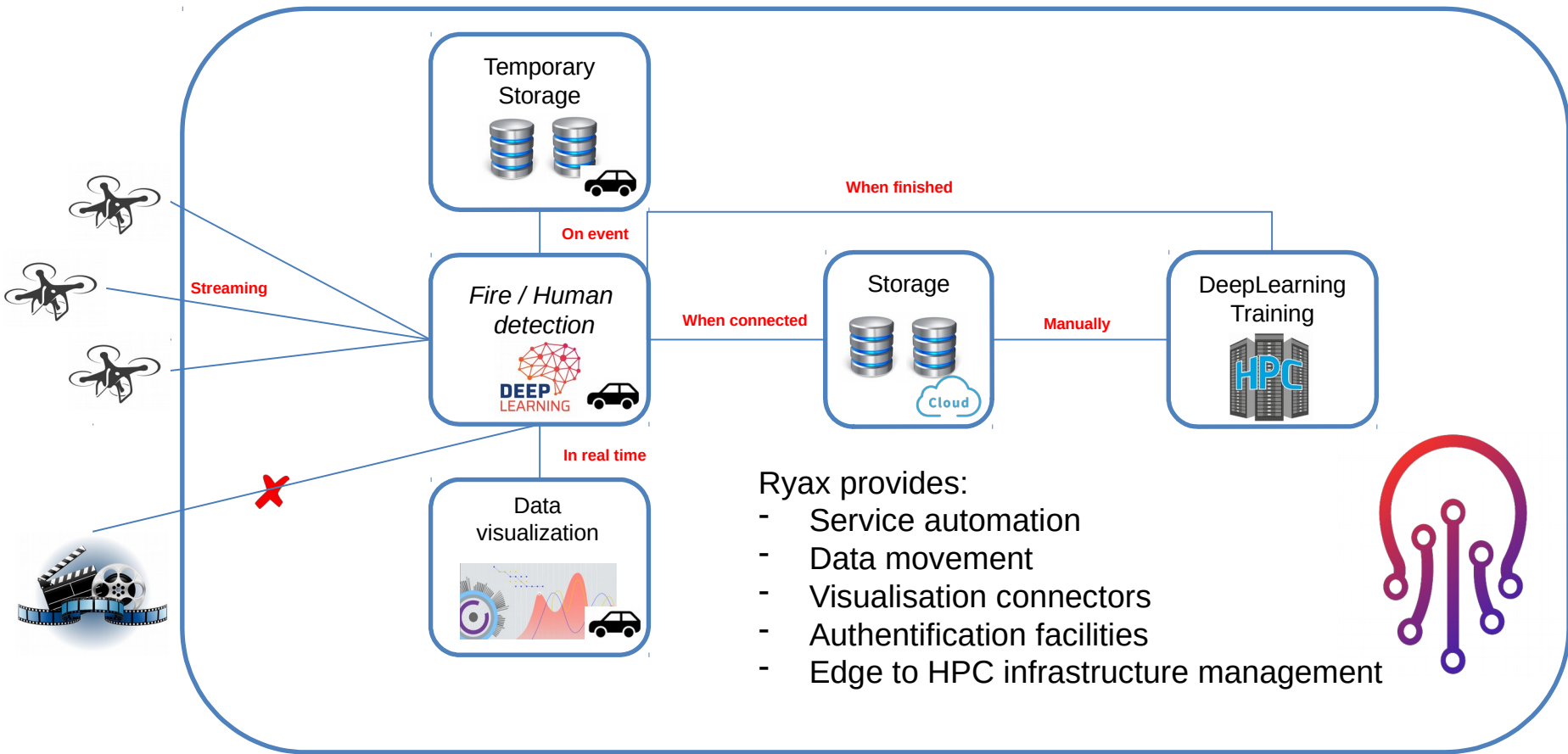
Transparent execution upon every infrastructure (On premise, Edge, Cloud)



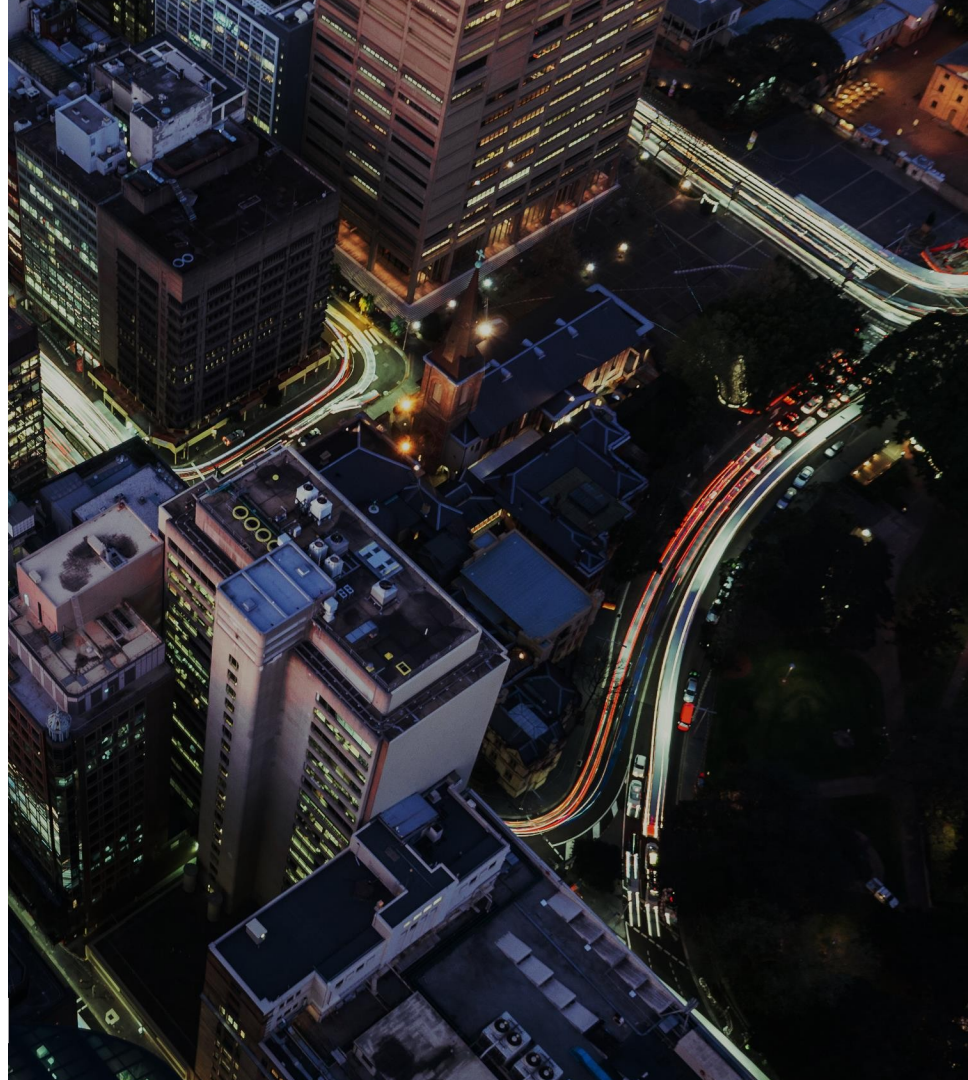
# Use Case: Predictive Maintenance



# Use Case: Fire detection by drones



# Ryax internals and R&D



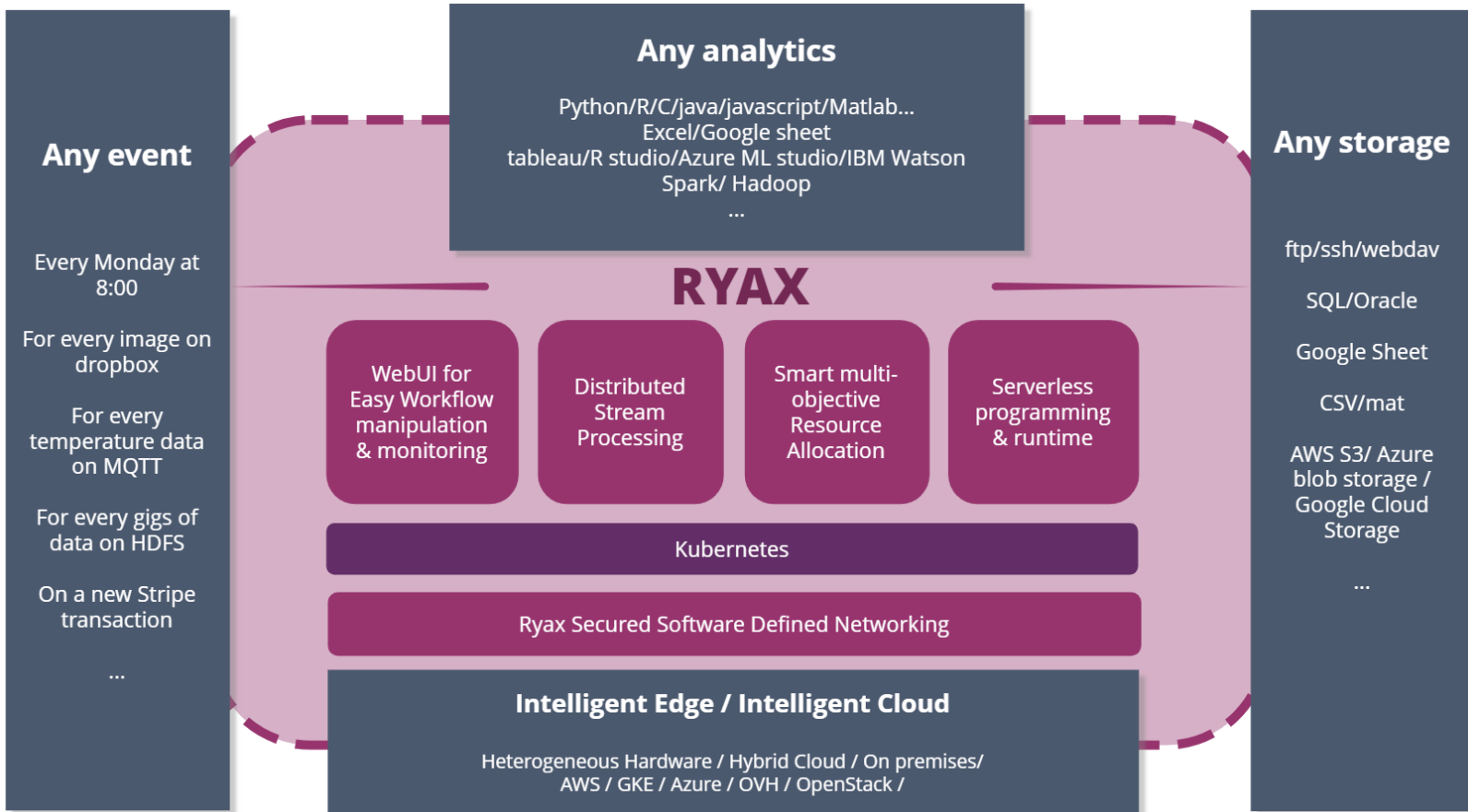
# Data engineering for Near Real Time

## ***Data Engineering challenges for hybrid edge-cloud infrastructures:***

- Network and Hardware Heterogeneity
- Seamless Programming and Deploying streaming applications
- Multi-level, multi-user data privacy
- Isolation, security, billing
- Content-aware, intermittent connections, locality, scalability



# Ryax High Level Architecture





# Ryax - SDN layer

## ➤ **Software Defined Networking** management side :

- **VPN creation** so that workflows can run upon a secure encrypted private network
  - A virtual network interface is created and an IP is given that remains the same no matter what type of network we use underneath (Wifi, Ethernet, 3G/4G)
  - Transparent and dynamic migration between networks
- Working towards providing mechanisms for **authentication** of new compute devices on a VPN
  - This will serve in allowing the edge gateways of each bus or drone to communicate amongst them and with the primary server(s) using a VPN that can be dynamically controlled
  - Manual authentication is implemented and currently working on more automated strategies

# Ryax - Orchestration layer

- Offering Serverless Computing based upon Kubernetes orchestration :
  - Function deployment currently based on Docker but implementing lower overhead technology leveraging cgroups
    - Evaluating Kata-Containers, gvisor, Firecracker
    - Working in minimizing overhead and optimising bin-packing of functions
  - Federation of hybrid edge-cloud clusters
    - Currently connected on one Kubernetes cluster based on SDN
    - Evaluating initiatives such as Federation v2 of k8s and software such as microK8s, k3s for low-overhead K8s at the edge
  - Scheduling optimization based on multi-criteria/multi-objective techniques



# Ryax R&D - EU Funding

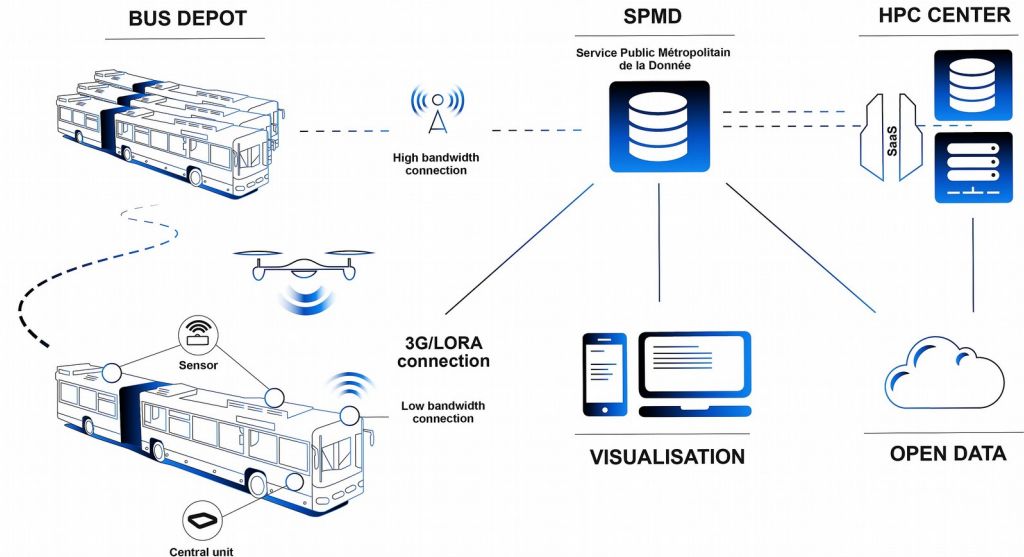
- Funded Project CEF named AQMO upon “**Air Quality and Mobility**” , HPC analytics for air quality on transportation vehicles, in collaboration with: Irisa, Rennes Metropole, GENCI/Idris, etc
- Funded project H2020 **ICT-11a** : “Large-scale HPC-enabled industrial pilot test-beds supporting big data applications” named Cybele with subject: **Fostering precision agriculture and livestock farming through secure access to large scale HPC-enabled virtual industrial experimentation environment empowering scalable big data analytics** in collaboration with Ubitech, Cineca, ICCS, WIT, CERTH, BSC, Bull/Atos, etc

# Ryax - AQMO project

AQMO project <http://aqmo.irisa.fr/> will offer an end-to-end urban platform that extends current practices in air quality measurements using innovative services based on computing simulation.

→ Ryax within AQMO is responsible for data analytics workflow tasks automation and orchestration from Edge gateways up to HPC clusters. R&D to allow:

- 1) Deployment tools overhead minimization for adaptation at edge
- 2) Orchestration in a mobile environment and dealing with intermittent connectivity
- 3) Resource allocation optimizations through new multi-criteria scheduling algorithms



# Ryax - Cybele project

Cybele project <https://cordis.europa.eu/project/rcn/220361/factsheet/en> develops large scale HPC-enabled test beds and delivers a distributed big data manager architecture and a data management strategy for improvements in precision agriculture and livestock farming

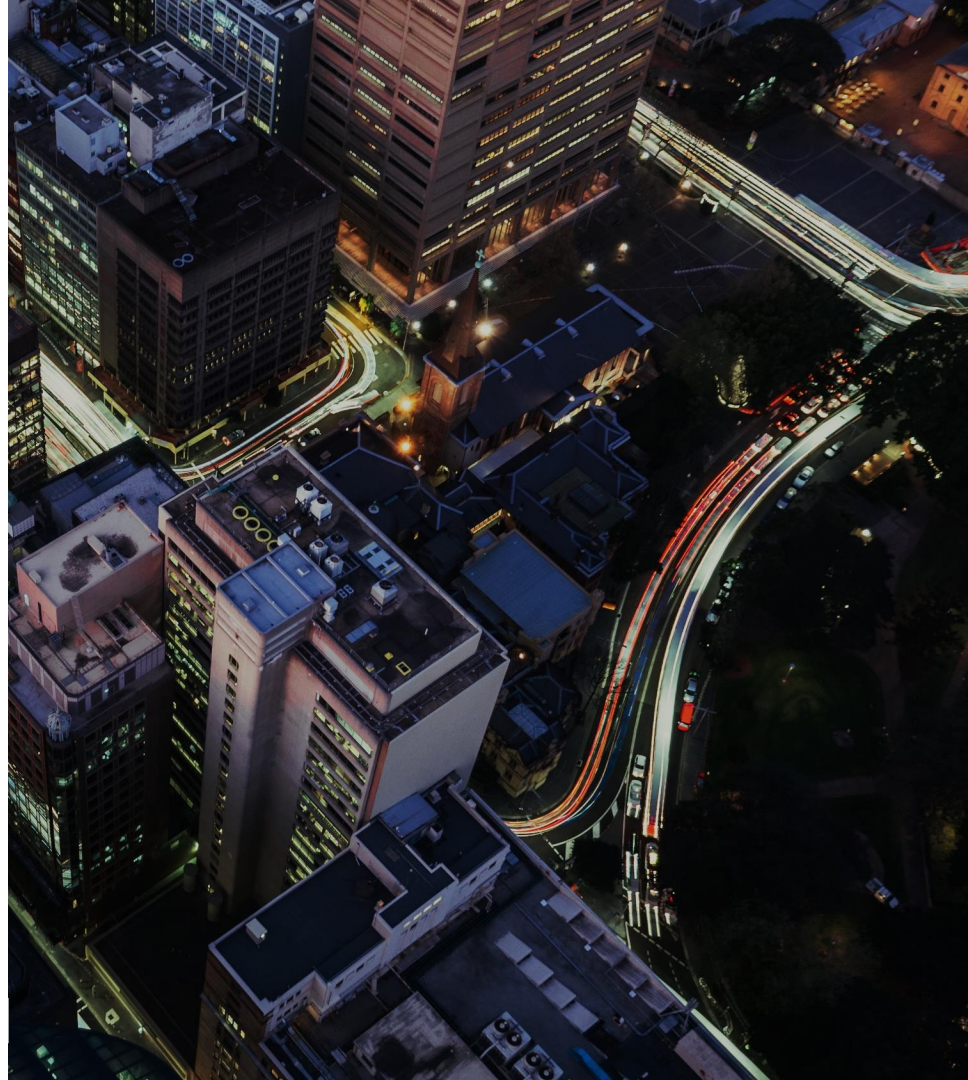


→ Ryax within Cybele is responsible for:

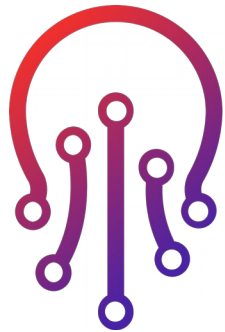
- 1) The data analytics workflows task automation
- 2) The integration of Kubernetes as the **orchestration engine** and development of mechanisms to support its communication **with different HPC resource managers and Big Data** frameworks.
- 3) Development of **Scheduling** algorithms to improve the **co-execution of mixed HPC and Big Data** workloads within heterogeneous computing facilities.
- 4) The study and extension of **simulation software** Batsim, which focuses on batch processing job scheduling, to better **consider the specificities of Big Data** workloads and experimentation with new scheduling algorithms for mixed workloads.



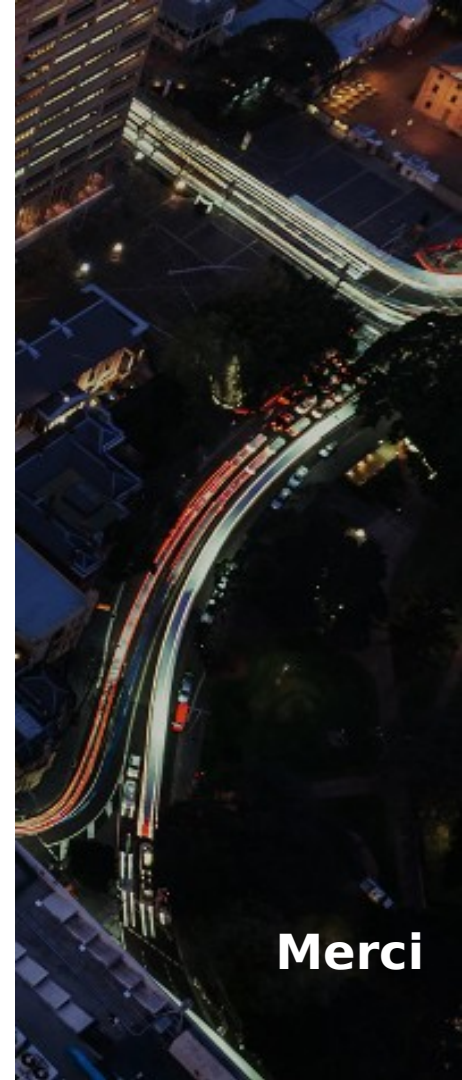
# Demo Time



*Yiannis.Georgiou@Ryax-Technologies.com*



**ryax**  
technologies



**Merci**