



# HPC with Hybrid Clouds for Critical Big Compute workloads

by Denis Caromel, CEO & Founder, ActiveEon



# Company



## Key information

- ISV Founded in 2007 by Denis Caromel in Sophia-Antipolis, Spin-off of INRIA
- **400 Man-Year R&D Investment**
- 60% of the revenue from international

## ProActive Solution

Scheduling  
Orchestration  
Meta-scheduling  
Resource Allocation

On-premises and on All Clouds  
Open Source

## Partnerships

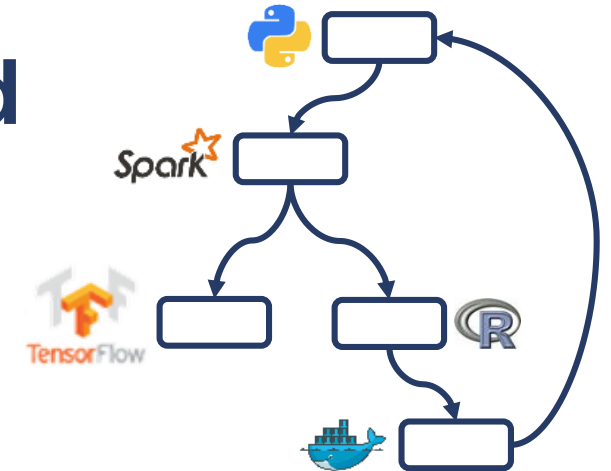


## Global Locations



- Sophia-Antipolis (France)
- Paris (France)
- London (United Kingdom)
- San-Jose (United States)
- Montreal (Canada)
- Fribourg (Switzerland)
- Dakar (Senegal)

## Build and Deliver to the Market the Best Workflows/Orchestrator Suite for Automation & Big Compute\* in the Cloud



At the core of the Revolution:

Big Data – AI / Machine Learning – IoT – Workload Automation – Cloud Migration

Aka HPC Big Data

\* *Big Compute: Business applications in need of high number of Cloud VMs/Containers*

# AE Mission & Vision

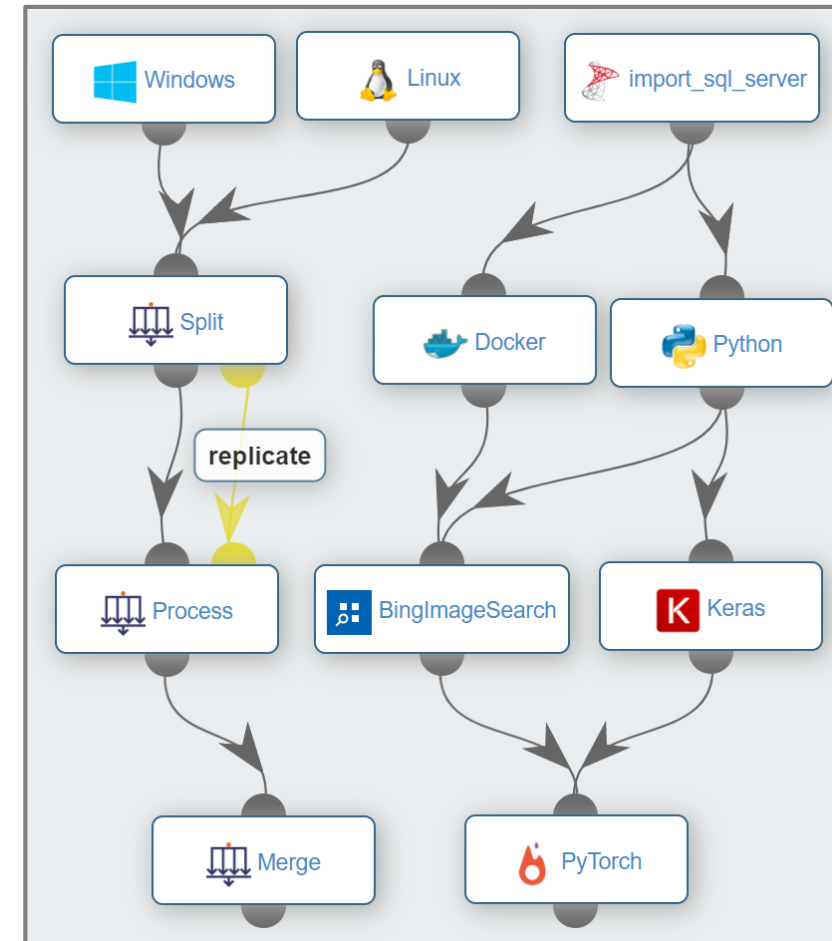
## Orchestrator Suite for Automation & Big Compute in the Cloud

Workflows = Automation

For Application People,  
Data Scientists, ML + IA Experts

Resource Management

For Hybrid and Effectiveness on the Cloud

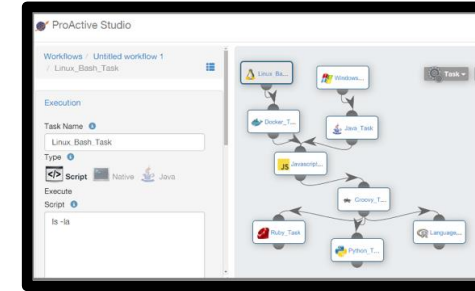


# Next Generation Scheduler/Orchestration **Activeeon** SCALE BEYOND LIMITS

</> Open REST API

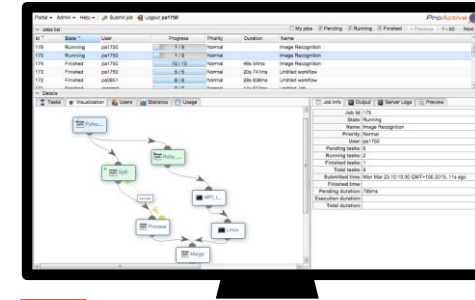
## Processing and Automation Workflows

<p>Any language</p> 	<p>Secured Data Transfers</p> 	<p>Meta-scheduler</p> 	<p>ETL, ERP, ELT, ...</p> 	<p>Full integration</p> 	<p>Translator</p> 
---	---	---	--	---	---







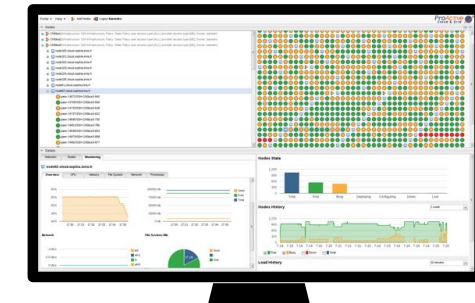
## Scheduler and Orchestration

<p>Priority &amp; Planning</p> 	<p>Parallel Executions</p> 	<p>Error Management</p> 	<p>Multi Users</p> 
--	--	---	--



## Resource Management and Monitoring

<p>Multi-platform</p> 	<p>Local Machine</p> 	<p>Network Resource</p> 	<p>Batch Scheduler</p> <p>Slurm PBS SGE LSF</p>	<p>Cloud</p> 
---	--	---	---	--



# Fault-Tolerance & Error Management

Id	State	Issues	User	Progress	Priority	Duration	Name	Project
2	Paused	Pause			Normal		Untitled workflow 1	Not Assigned
1	Finished	Restart All In-Error Tasks Resume All Paused Tasks Resume All Paused Tasks & Restart All In-Error Tasks			Normal	20s 910ms	Untitled workflow 1	Not Assigned

< First < Previous

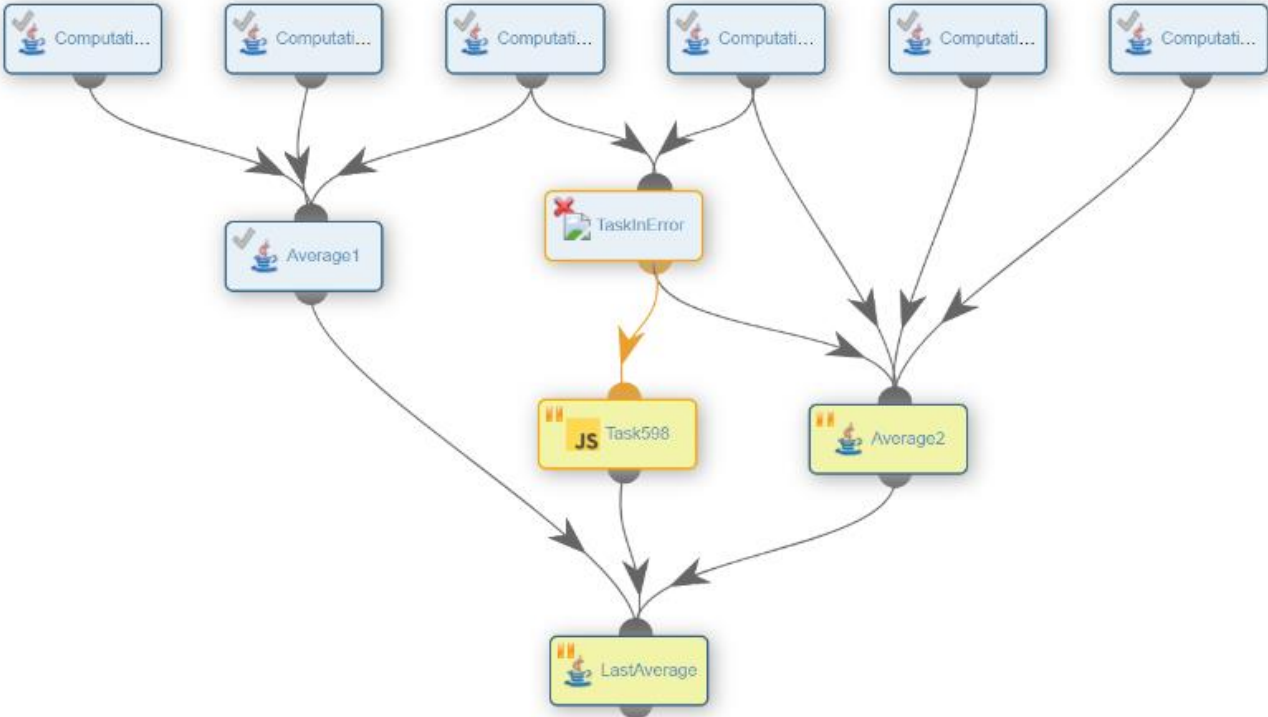
Details

Tasks Visualization

Priority

Remove

Kill



# Cloud Automation: On-demand PaaS



ProActive Automation Dashboard

Help ▾

Automation Dashboard Workflow Studio Scheduling & Orchestration Resource Manager caromel

- Workflow Automation
- Workflow Catalog
- Job Planner
- Cloud Automation
- Notification Service 8

9  
Services in catalog

4  
Activated Services

2  
Transitioning Services

5  
Activated Services Today

1  
Finished Services Today

4  
PCA Running Jobs

6  
PCA Total Jobs Today

### Activated Services

Instance ID	Service ID	Info	Current state	Actions (Workflow name)	Endpoint list	Kill
2	Visdom	<span style="color: blue;">i</span>	RUNNING	Finish_Visdom ▶	visdom-server-1 <a href="#">↗</a>	<span style="color: red;">✕</span>
3	Zookeeper	<span style="color: blue;">i</span>	RUNNING	Finish_Zookeeper ▶	zookeeper-server-1 <a href="#">↗</a>	<span style="color: red;">✕</span>
4	Kafka	<span style="color: blue;">i</span>	VOID → RUNNING ⚡	No possible action	No endpoint	<span style="color: red;">✕</span>
5	Storm	<span style="color: blue;">i</span>	VOID → RUNNING ⚡	No possible action	No endpoint	<span style="color: red;">✕</span>

### Services and workflows

Service Activation | Full Services View | Service Jobs

Storm

Storm  
VOID→RUNNING

ProActive

PCA\_example\_sta  
rt  
VOID→RUNNING

HDFS-Spark

reserve\_nodes  
VOID→PA\_NODES\_  
RESERVED

Zookeeper

Zookeeper  
VOID→RUNNING

Kafka

Kafka  
VOID→RUNNING

Visdom

Visdom  
VOID→RUNNING

### Finished Services

Instance ID	Service ID	Info	History	Current state	Endpoint list	Clean
1	ProActive	<span style="color: blue;">i</span>	<span style="color: blue;">↺</span>	FINISHED	documentation <a href="#">↗</a>	<span style="color: blue;">🗑️</span>

On-Demand PaaS Services with full Life-Cycle Management



# 2. Scalability, Resilience, Elasticity

---

20 000 Cores on Azure

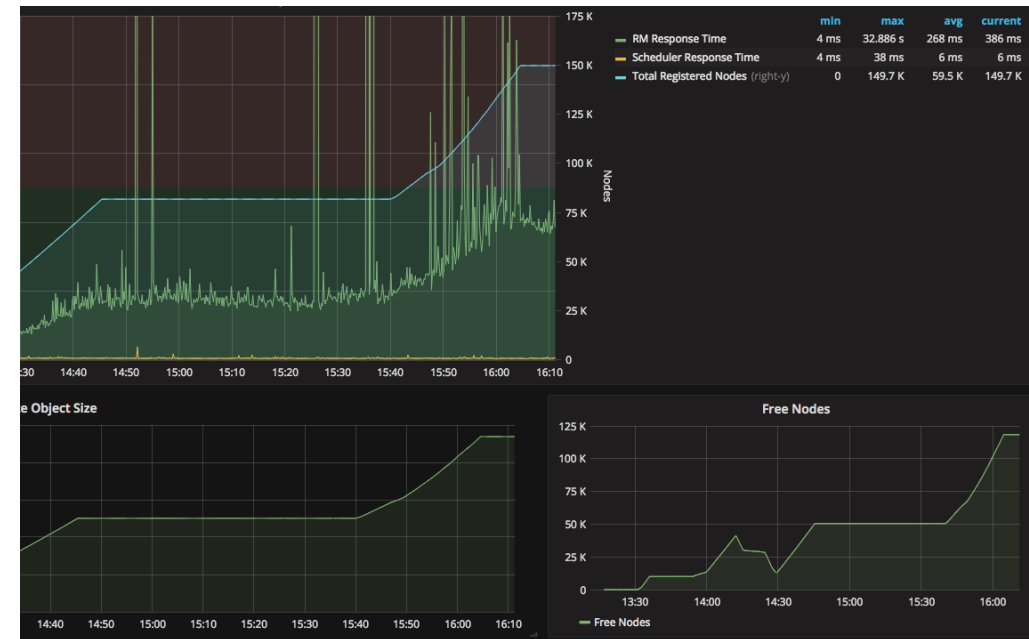




# 20 000 Cores Azure Benchmarks

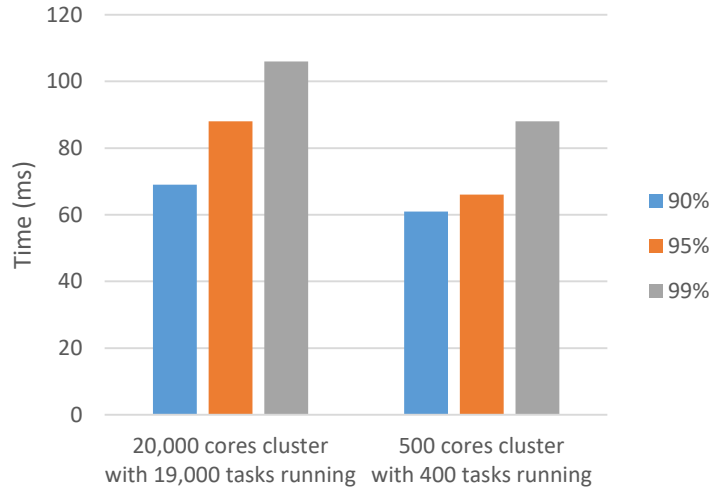
With ActiveEon Workflows & Scheduler:

- 15 mn to trigger and acquire 20 000 Azure Cores and to schedule 20 000 Tasks
- 99% of requests having less than 90ms response-time
- On 20K Cores, with 19K running Tasks, only 5 sec. to detect a software failure and redeploy the Task
- With 19K running tasks, only 30.8 sec. to execute a Job with 10 Tasks of each 30 sec., 97.4% efficiency.

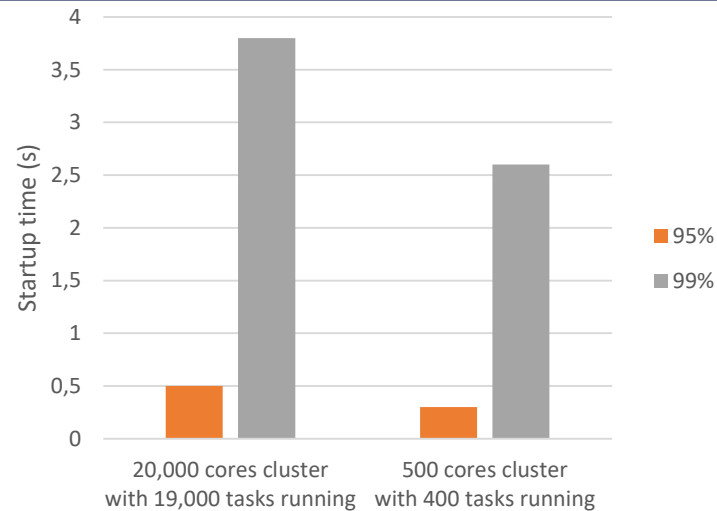


# Benchmarks on Azure Datacenter

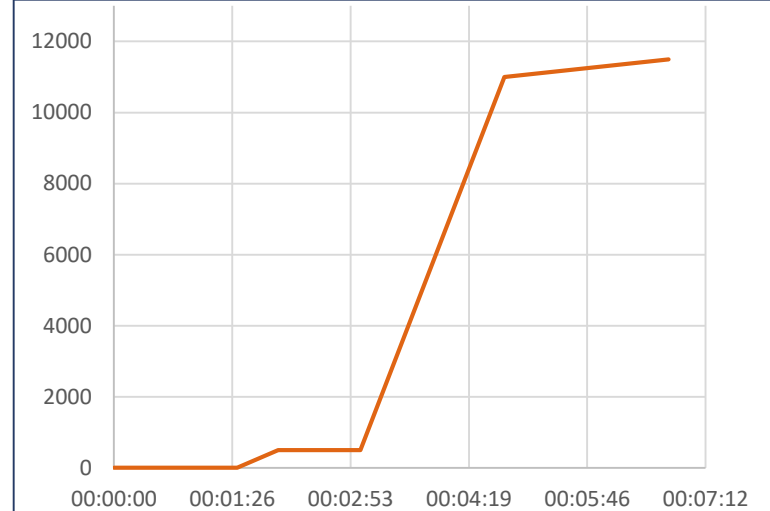
## Responsiveness < 110 ms



## Startup Time < 4 sec



## 10k Cores Available in < 8 min

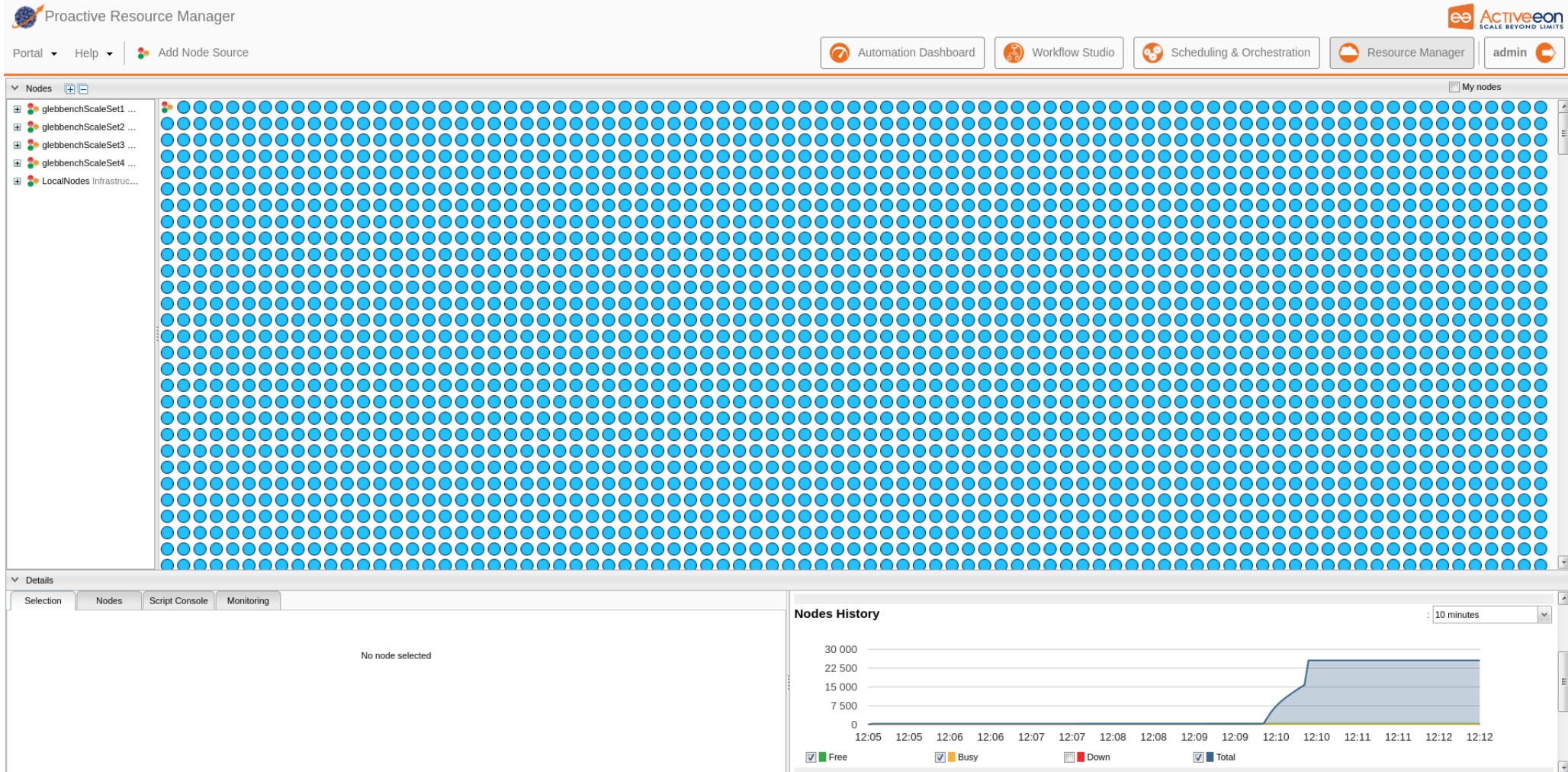


Upon failure acknowledgment,  
it takes 5 sec to redeploy the task on another core!

SLOs achieved on **20,000 cores utilized at 95%**

# 20 000 Cores Azure Benchmarks

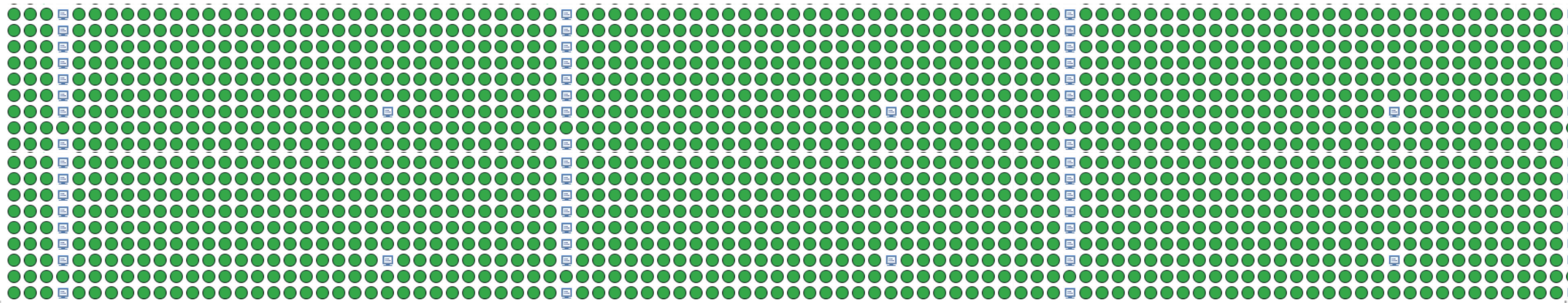
➤ Triggering 24 000 Azure VMs in a minute





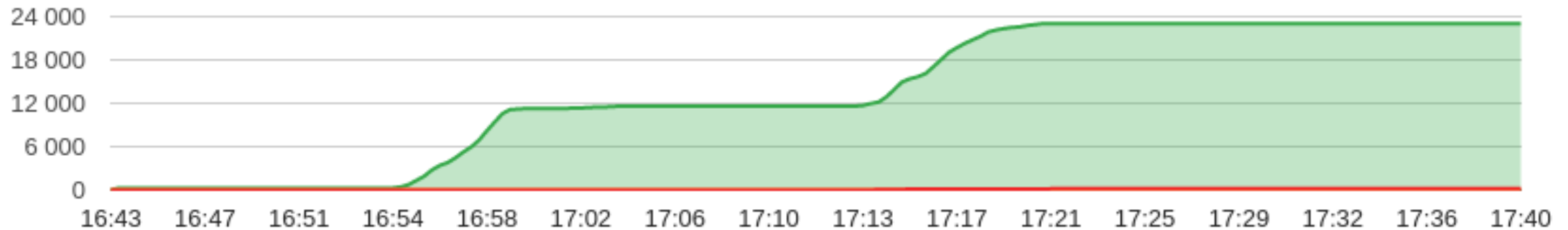
# 20 000 Cores Azure Benchmarks

➤ Acquiring 24 000 Azure Cores, here in 2 steps of 4mn each



Nodes History

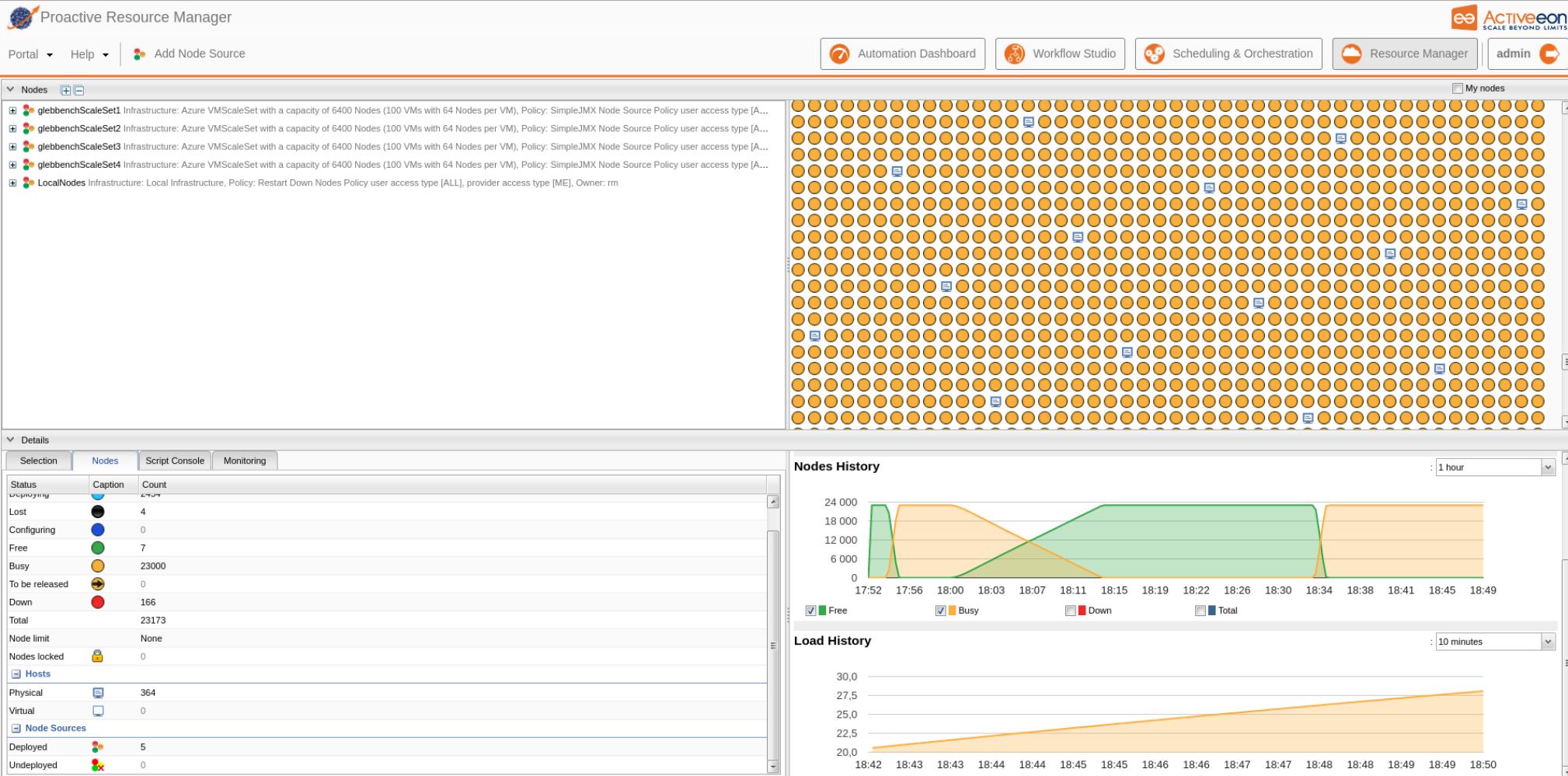
: 1 hour



Free  Busy  Down  Total

# 20 000 Cores Azure Benchmarks

## ▶ Deploying 24 000 Tasks



The screenshot displays the Proactive Resource Manager interface. At the top, there are navigation tabs for Automation Dashboard, Workflow Studio, Scheduling & Orchestration, Resource Manager, and an admin user profile. The main area is divided into two sections: 'Nodes' and 'Details'.

**Nodes Section:** A list of node sets is shown on the left, including 'glebenchScaleSet1' through 'LocalNodes'. On the right, a large grid of orange circles represents the deployed nodes, with a 'My nodes' checkbox at the top right.

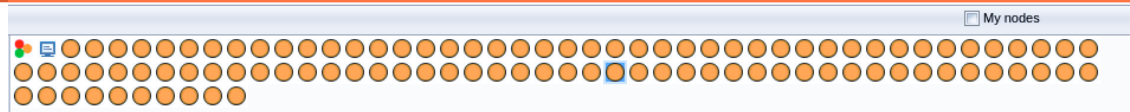
**Details Section:** A table provides a summary of node statuses and counts. Below the table, there are two charts: 'Nodes History' and 'Load History'.

Status	Count
Lost	4
Configuring	0
Free	7
Busy	23000
To be released	0
Down	166
Total	23173
Node limit	None
Nodes locked	0
Hosts	
Physical	364
Virtual	0
Node Sources	
Deployed	5
Undeployed	0

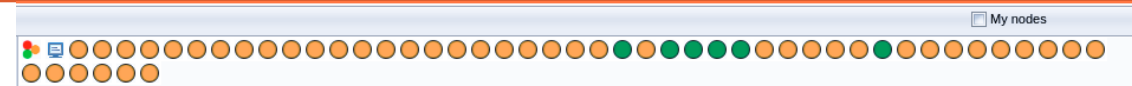
**Nodes History Chart:** Shows the number of nodes in different states (Free, Busy, Down, Total) over time from 17:52 to 18:49. The y-axis ranges from 0 to 24,000. The 'Busy' state (orange) peaks at approximately 23,000 nodes around 18:00. The 'Free' state (green) drops to near zero during this period.

**Load History Chart:** Shows the load over time from 18:42 to 18:50. The y-axis ranges from 20.0 to 30.0. The load starts at approximately 20.0 and increases steadily to about 27.5 by 18:50.

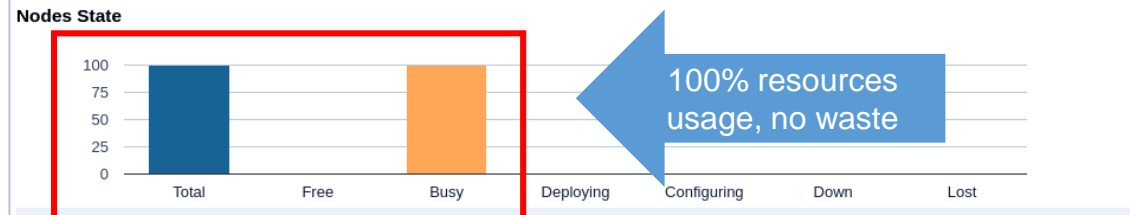
# Elasticity: Automatic On-Demand VMs



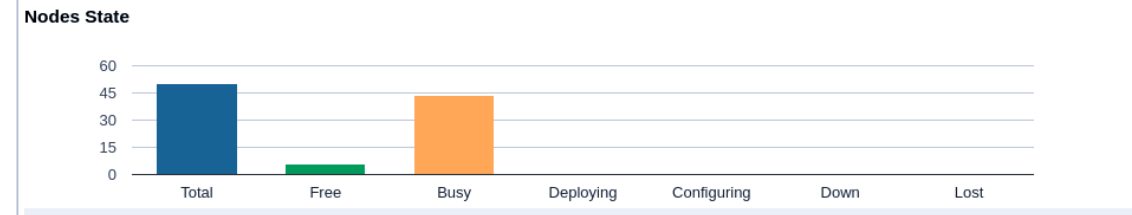
Provides cloud computing power according to your needs. Minimize costs by deploying VMs only when needed (configurable load factor). Never exceed your budget (min/max VMs threshold).



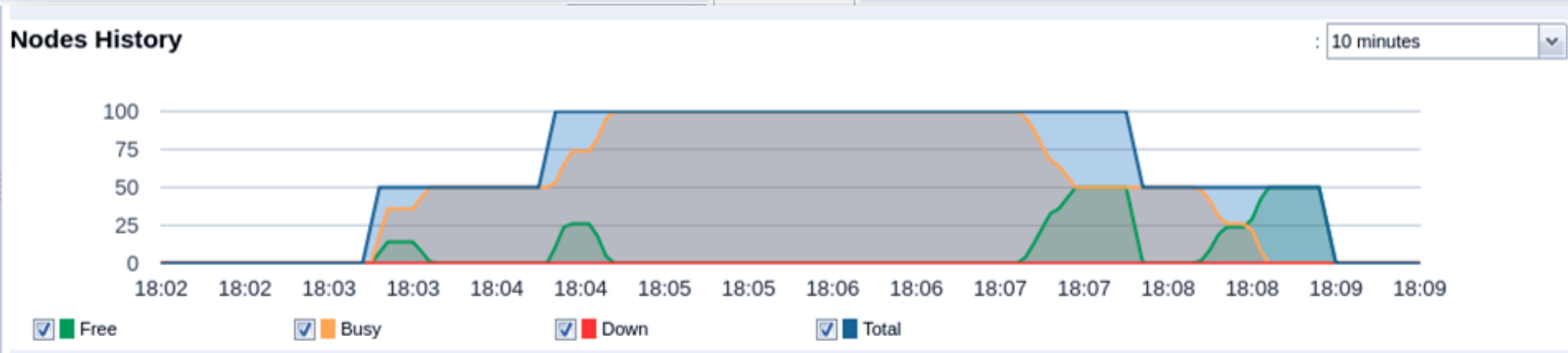
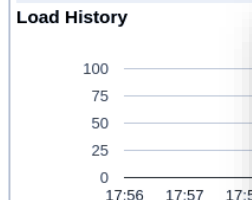
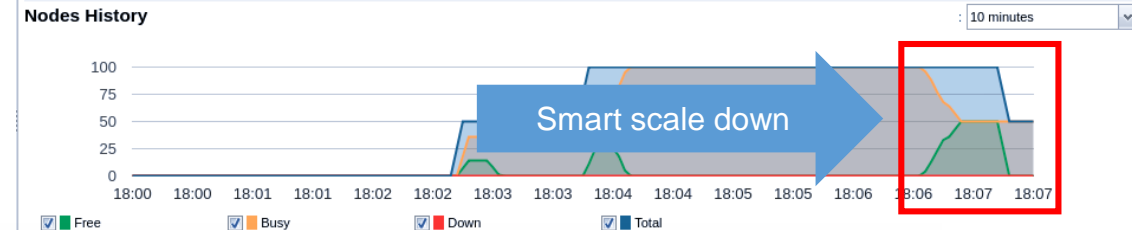
Smart and fully configurable elastic policy. Shutdown unused VMs whenever it's possible. Prevent time-consuming re-deployments by adjusting idle nodes' release delay (avoid scale up/down cycles).



100% resources usage, no waste



Smart scale down





## 2. Some Typical Customer Cases:

---

Large Worldwide International Companies

Early Adopters

Using ActiveEon for Critical Business Applications



Finance

IoT

Gov.

Manufacturing

Automotive

Aerospace

Nuclear

RedHat OpenShift

# PEPS Project

by  cnes







# PEPS: Plateforme d'Exploitation des Produits Sentinel

Redistribute for free the products of Sentinel satellites, S1A, S1B, S2A and S2B, S3A and S3B from COPERNICUS, the European system for the Earth monitoring.

Multi-sensor (radar, optical, etc.), High frequency, long term project.

1 PB in 20 years and 7 PB in 2 years! 10 TB/day



# Space & Image Processing



## Situation

Make Sentinel data available to the greatest number and encourage the development of applications using them (agriculture, maritime field...)

## Solution

Proactive Solution provided by ActiveEon to execute on Azure in hybrid mode allows enhancing PEPS data and making them available to API providers :

- Multi-Cloud Ecosystem Platform
- Remove complexity for Data Scientists
- Provide Cloud performance

## Benefits

- Faster execution, Optimisation of On-Prem ressources & Clouds,
- Easier to use by end-users







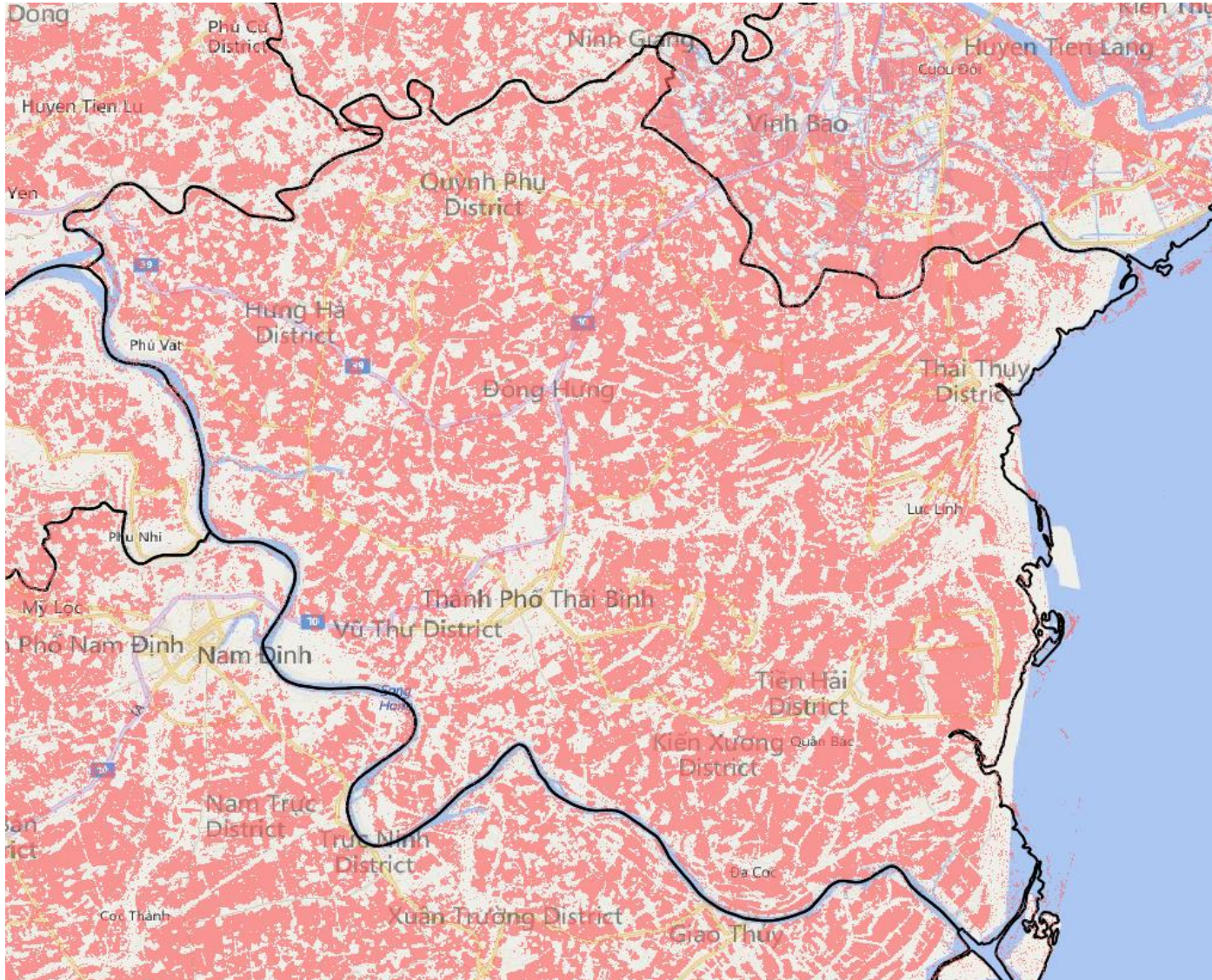
## Key points to choose AE:

- OnPrem & Clouds with a unique Task definition
- Single Interface for Job Management
- Out-of-the-box effective error management
- Product integration with Docker
- Business tasks can trigger VM provisioning on Azure Cloud





# Example of Result: Rice Growth and Harvest Prediction



**Winter-Spring Rice Map,  
Thai Binh province,  
Thailand**



# Finance

L&G a leading multinational finance and insurance company with headquarters in London



## Situation

European regulations: Solvency II, Basel III, etc.  
Transform legacy system and embrace Cloud computing

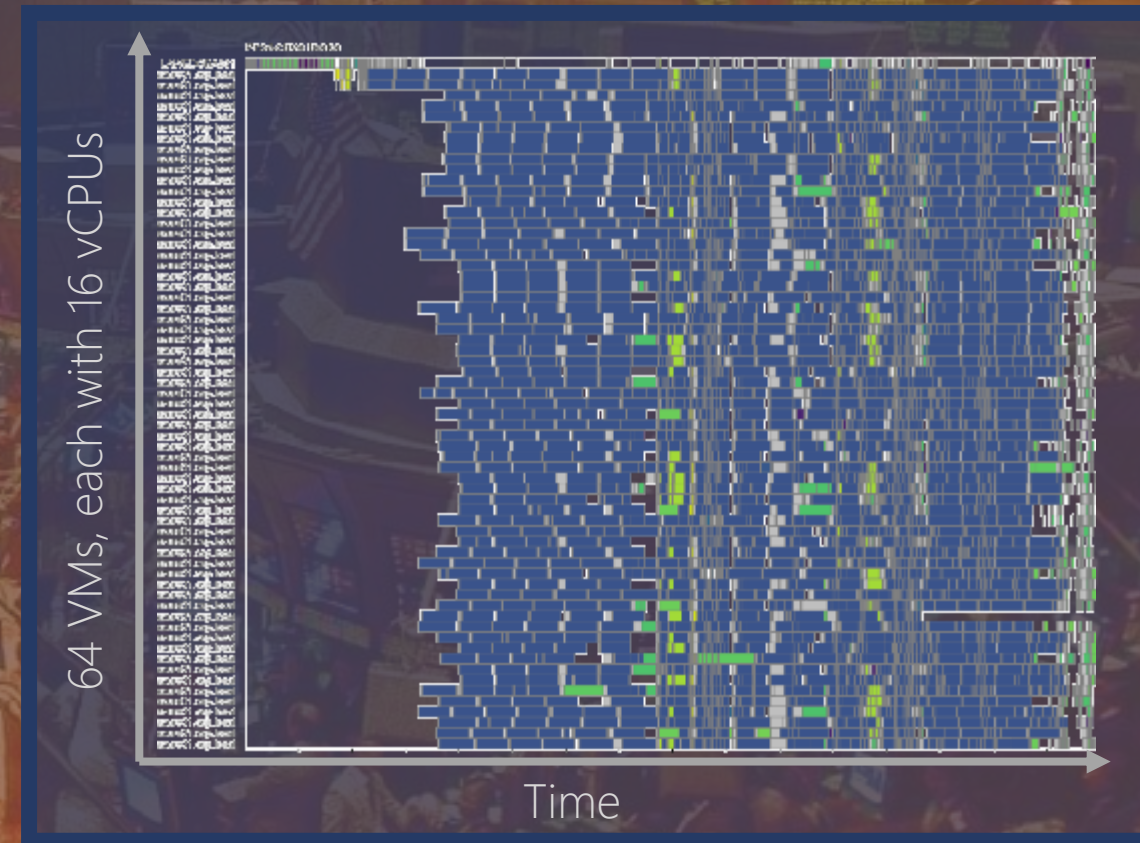
## Solution

Activeeon ProActive and migration to the Cloud have enabled faster and more reliable execution:

- Cloud bursting
- Error management
- Prioritization

## Benefits

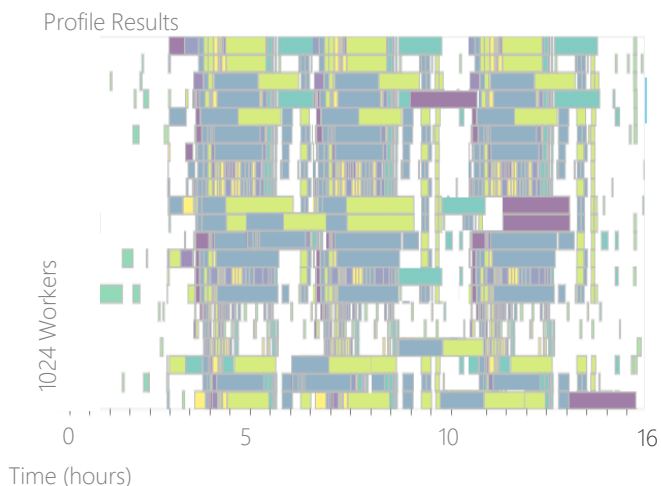
From 18 hours to 2 hours for priority reports  
Agile development with an objective of 4,000 cores



➔ \$1.2m / year committed spent on Cloud

## Profile Results *without* ProActive

32 VMs or Machines



Full computation without intermediate result

18H



Batch optimization with ProActive:

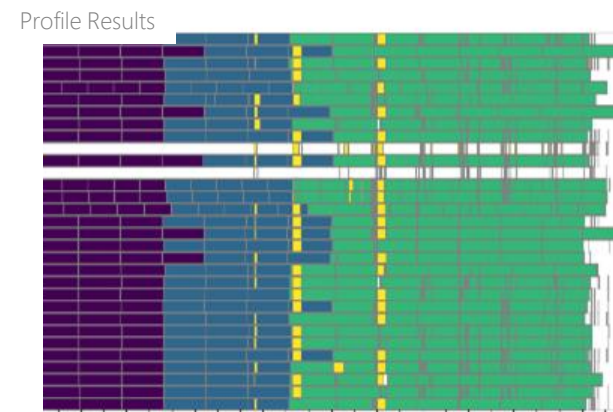
- Enforce strong priorities
- Optimal compact execution
- Start tasks as early as possible
- Pipeline and co-allocate
- CPU-intensive with I/O intensive tasks



## Profile Results *with* ProActive

*“More Value, Faster”*

32 VMs or Machines



High Priority Results

2H



Full Report

5H



32 VMs or Machines

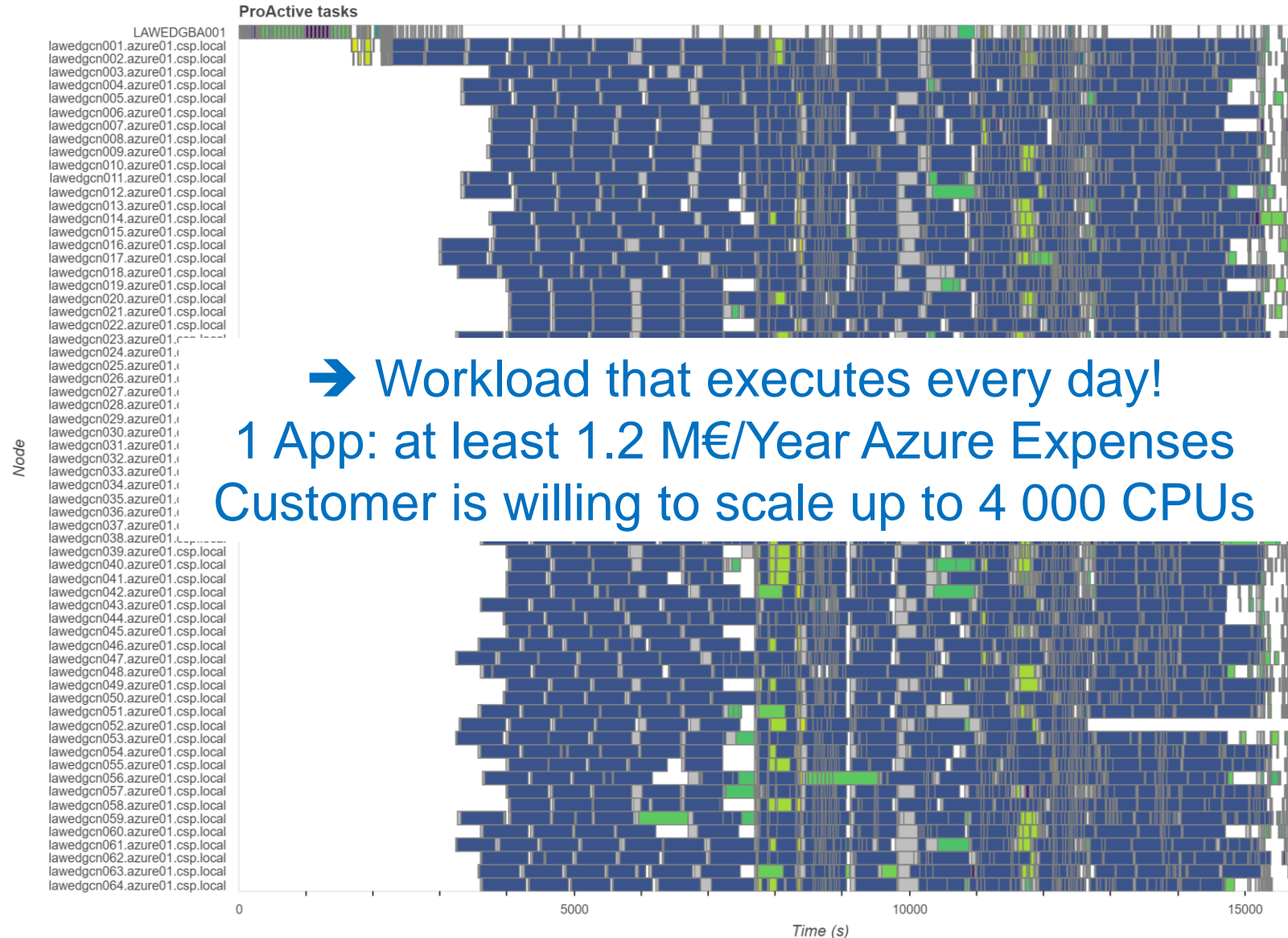
## Achievements & Benefits:

- High Priority Results in 2 H instead of 18 H !
- Compacting Executions for savings on the Cloud

**What Legal & General said about Batch optimization with ProActive:**  
*“It enforces strong priorities, optimally compacts execution, starts tasks as early as possible, pipelines and co-allocates CPU-intensive with I/O intensive tasks.”*



# Solvency : Executing on Azure, scaling to 1024 CPUs



➔ Workload that executes every day!  
1 App: at least 1.2 M€/Year Azure Expenses  
Customer is willing to scale up to 4 000 CPUs





# IoT



Komatsu is a Japanese multinational corporation.  
It manufactures construction, mining, industrial and military equipment.

## Situation

ActiveEon Orchestrates on Cloud execution over hot and cold storage for Streaming and Batch Analytics  
→ 1,200 tasks executed per hour

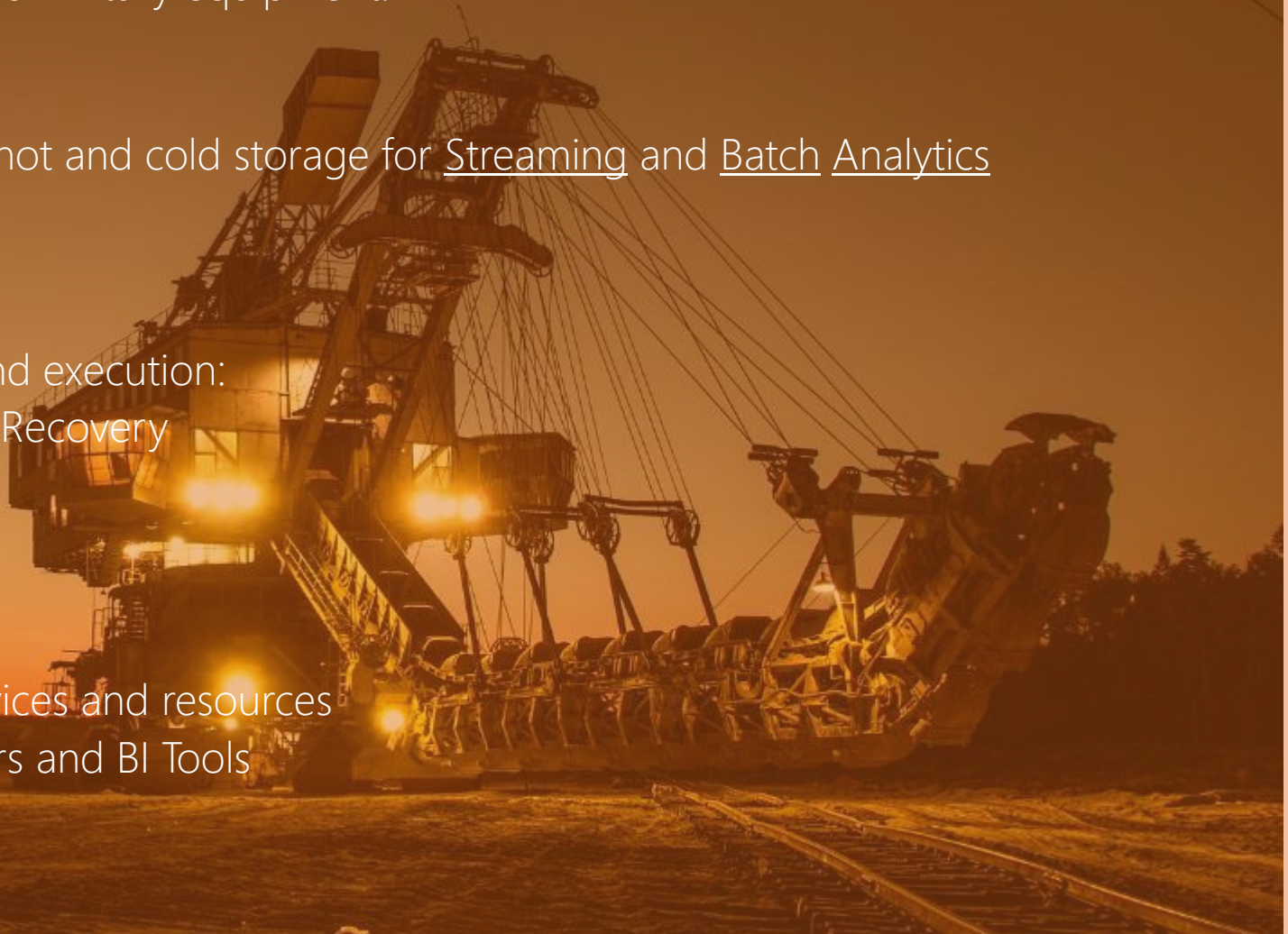
## Solution

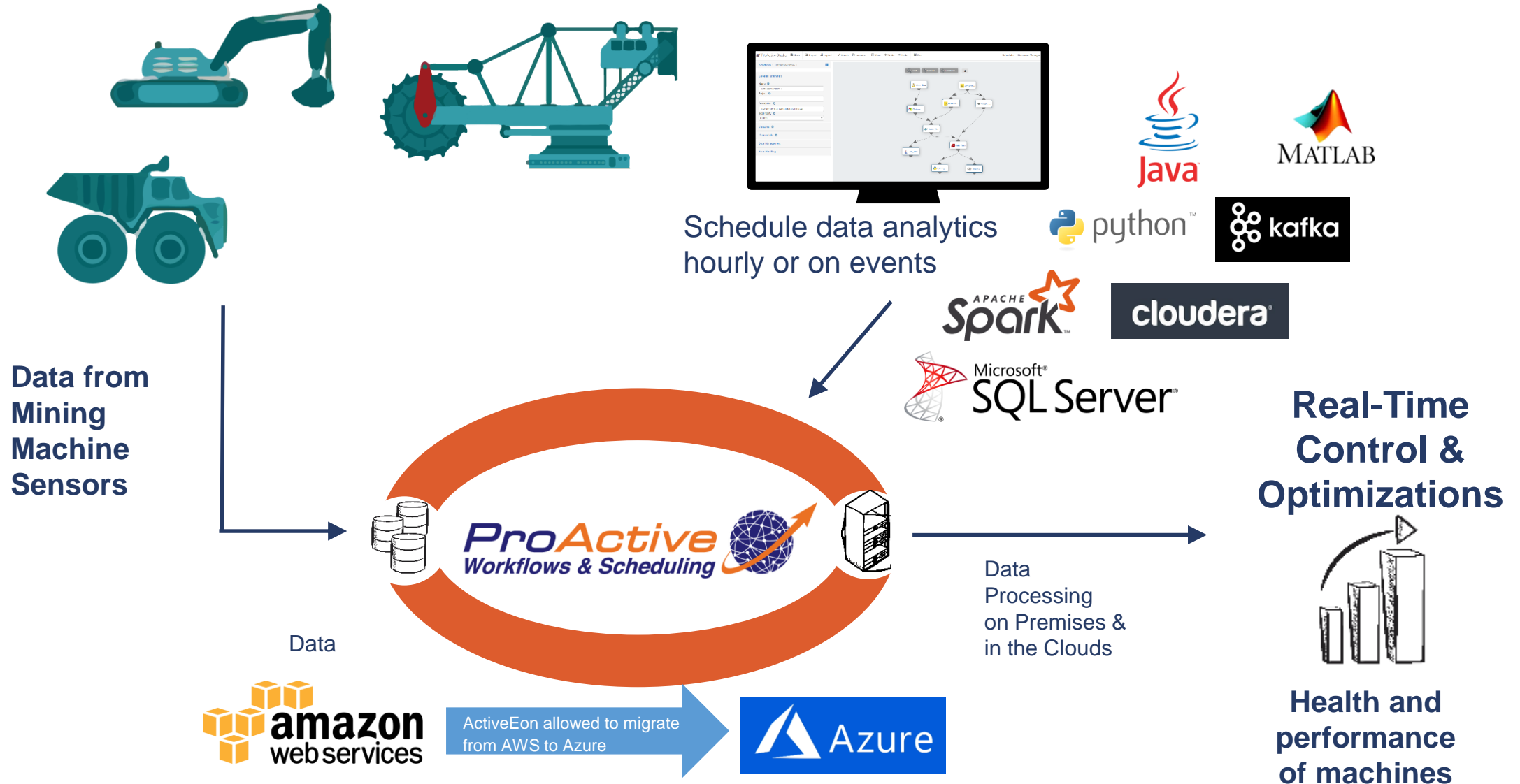
Activeeon has enabled control over scheduling and execution:

- Error Management – Notification, Automated Recovery
- Job Planner
- Distribution & Parallelization

## Benefits

- Reliable execution to orchestrate multiple services and resources
- Provide consistent results and KPIs to end users and BI Tools







# Gov.: UK Ministry of Interior



UK Ministry of Interior is using ActiveEon for 2 critical applications:

- Visa Delivery Process, and
- Big Data & Analytics platform for Crime Reduction (HODAC).

## Situation

25 different sources of Data.

Build a consolidated Data Lake and analytics platform to be used for many Home Land security applications.

## Solution

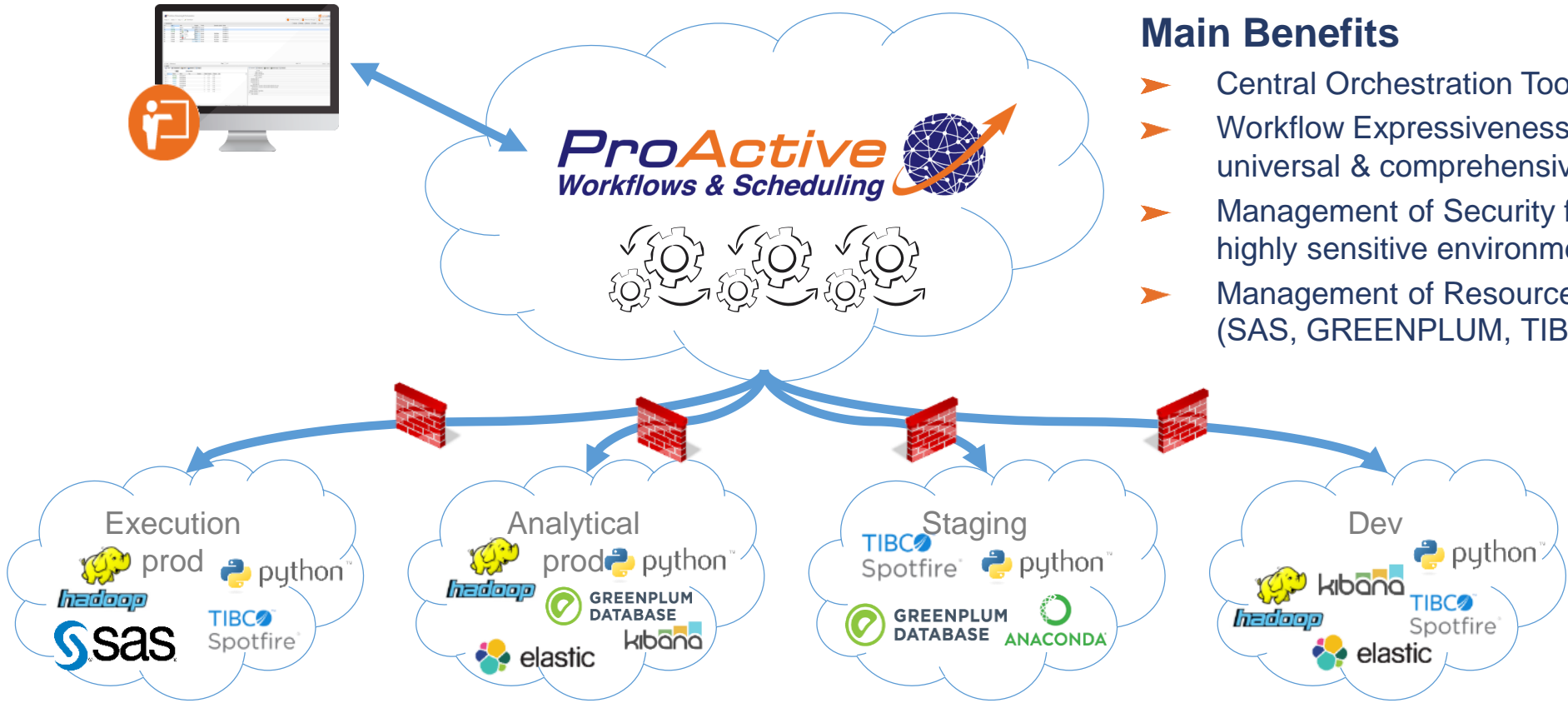
ActiveEon used as the central Orchestrator to Schedule and Meta-Schedule all the Big Data, ETL, Analytics, Machine Learnings software appliance of the platform (Hadoop, SAS, TIBCO Spotfire, Python, Anaconda, GreenPlum, ElasticSearch, ...).

UK Border

# Meta-Scheduler, Big Data ETL + new ELT



« The only solution capable to Schedule any Big Data Analytics, mono-threaded, multi-threaded, multi-core, parallel and distributed »  
 Cap Gemini Lead Engineer for Home Office

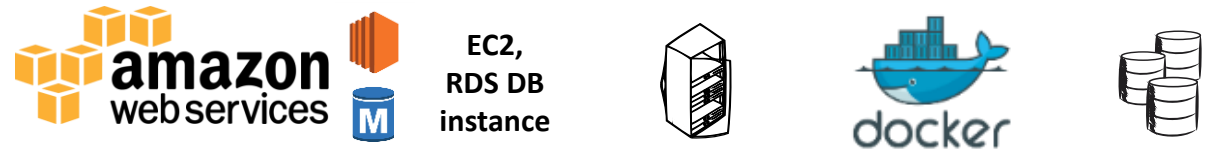


## Main Benefits

- Central Orchestration Tool
- Workflow Expressiveness: universal & comprehensive
- Management of Security for highly sensitive environments
- Management of Resources for all appliances (SAS, GREENPLUM, TIBCO, ...)

## Virtualized Infrastructure using Docker

4 Thousands of physical cores





## COMPANY PROFILE

- Industry: BioTech
- Product: Metagenomics

## MAIN DRIVER

- Simple workflow process definition
- Workflow model and data management
- Compute migration from on-prem to the cloud

## REQUIREMENTS

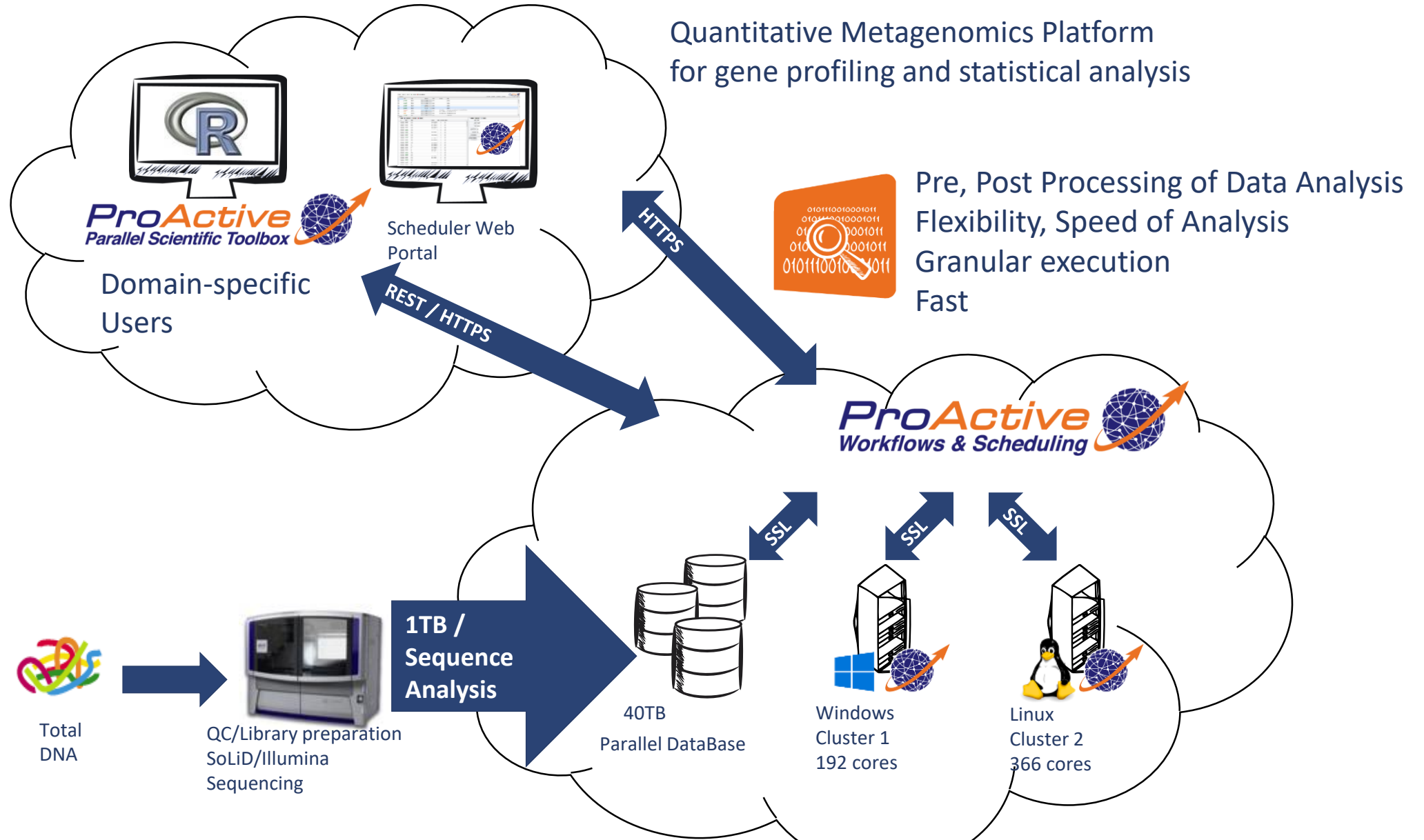
- Guidance and support to achieve high performances
- Fit in hybrid architecture multiplatform
- Integration with R
- FlexLM support (licenses manager)
- Remote Visualization for interactive tasks

## CHALLENGES

Process 500 terabytes per year  
Flexibility and enabler of interoperability  
between heterogeneous services  
Job affinity with data location  
Transfer sensitive data to the cloud for  
processing

## RESULTS

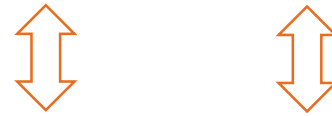
Efficient metagenomics pipeline  
Granular compute management  
User friendly system for maximum utilization  
Secure transfers



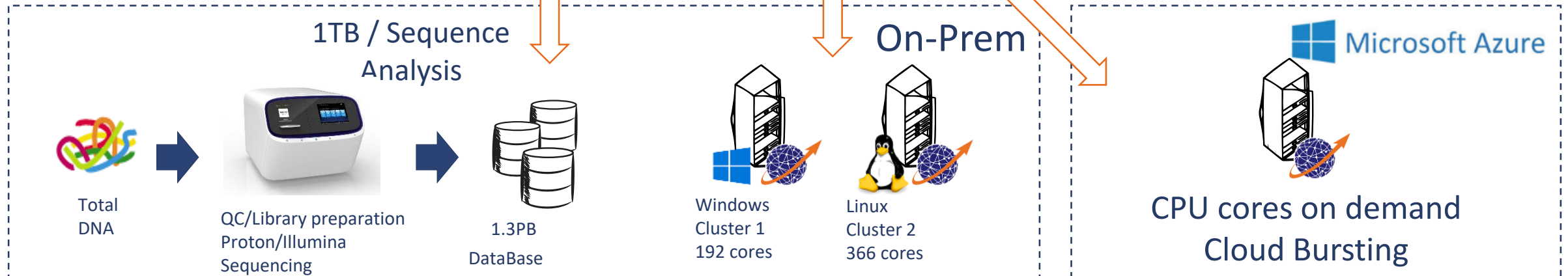
Quantitative Metagenomics Platform  
for gene profiling and statistical analysis



Web Portal and  
Integration with  
Scientific tools



Pre, Post Processing of Data Analysis  
Flexibility, Speed of Analysis  
Granular execution  
Distribution for fast execution  
Secure data transfer





## FEATURES

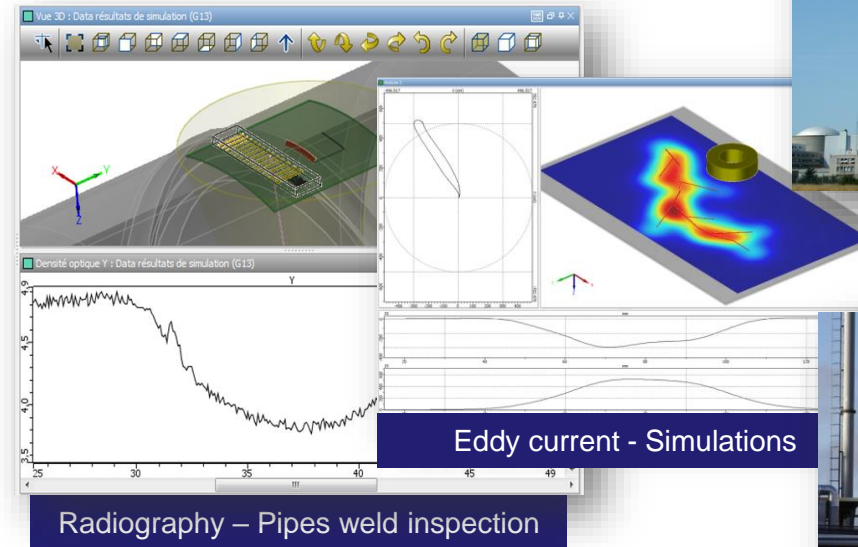
- NDE batch processing, parametric studies, non-regression tests on multiple clusters
- Transfer Input and Output data
- Event programming to follow executions
- Workflow process definition
- Activeeon guidance and support
- Cloud version: Execution on Microsoft Azure with 50 VMs/day per CIVA user → 25K nodes/year
- A potential of \$1M\$ Azure spending per Year

## BENEFITS

- Flexibility and enabler of interoperability between heterogeneous infrastructures
- Ability to run large POD (Probability of Detection) computations, which were taking months on a single computer
- Large-scale simulations with Microsoft Azure cloud

## ABOUT CIVA NDE SOLUTION:

- Multi-technique (Ultrasound, Eddy current, Radiography) software platform developed by the CEA LIST and its partners
- The software is distributed by EXTENDE and its distributors



## END USERS

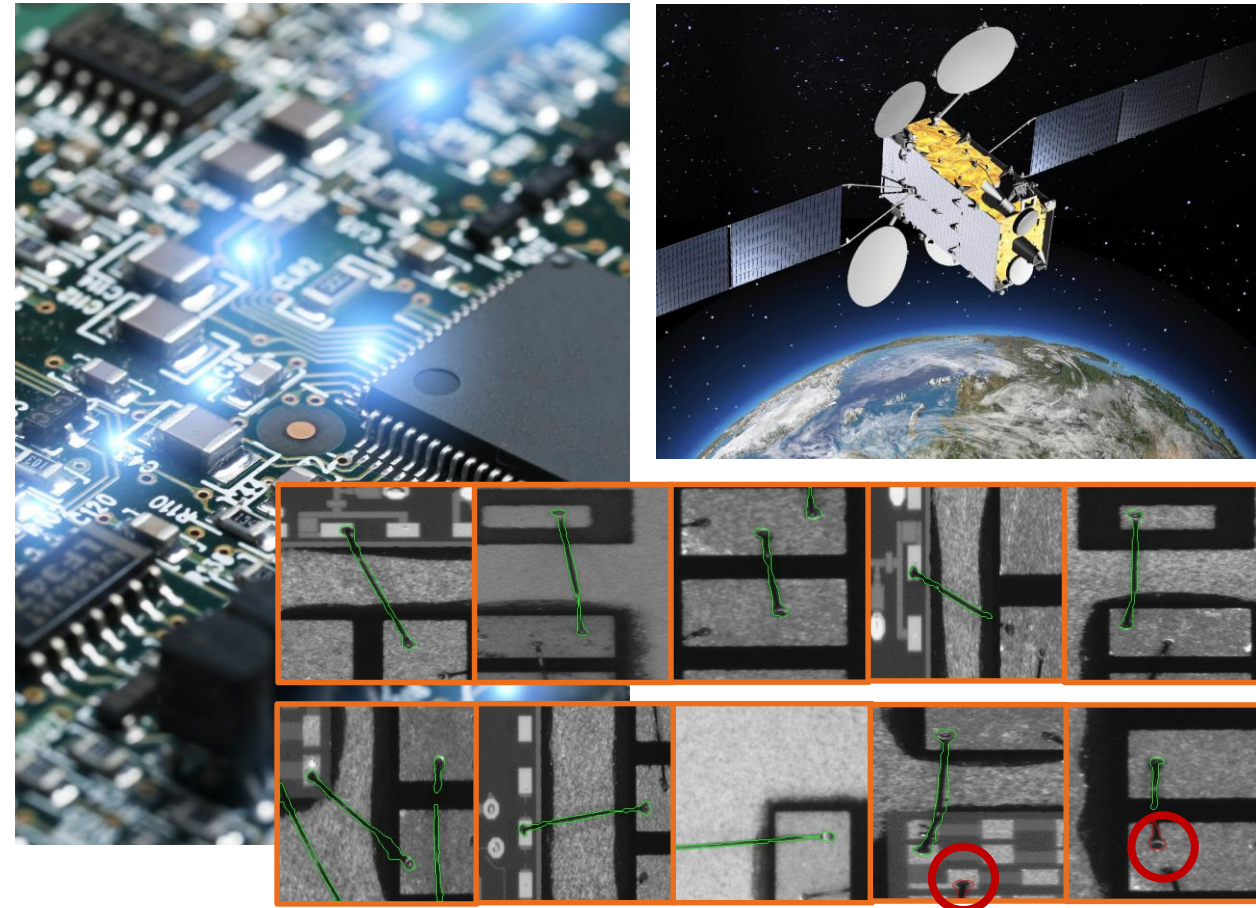
- Nuclear Energy, Oil & Gas, Aeronautics, Transportation

## FEATURES

- Detection of wires defect on a set of images from production line using Deep Learning
- Deep Learning on images of wires: occlusion, variation, noise, grayscale, semantic analysis
- Detection of defaults using a pre-defined wire model and computing a distance measure
- Workflows for model training and prediction for parallel execution

## BENEFITS

- Automatic detection of defaults in hybrid circuits manufacturing
- Higher precision of Machine Learning results
- Faster results with parallel execution of machine learning workflows
- Workflows can be used for other applications



Faulty wires come out in red





**Activeeon**  
SCALE BEYOND LIMITS

# Conclusion





# Customer Benefits:

- Same Workflows  
On-Premises & On Clouds
- Easier to use by end-users
- Faster Executions & Cost Reduction with  
Elastic Provisioning
- Scalability, Fault-Tolerance, Docker Support
- Optimized Deployment with ScaleSet from



**Azure**

# Free Download & Try Online: Try.ActiveEon.com

**ActiveEon**  
SCALE BEYOND LIMITS

**Denis CARMEL** PhD, MBA  
CEO & Founder  
denis.caromel@activeeon.com  
Phone +33 6 76 87 54 09

2000 route des Lucioles - Les Algorithmes  
06560 Sophia Antipolis, France

www.activeeon.com  
Phone +33 988 777 660

Open Source ISV & Professional Support



Paris, London, Sophia Antipolis, San Jose, Montreal, Dakar