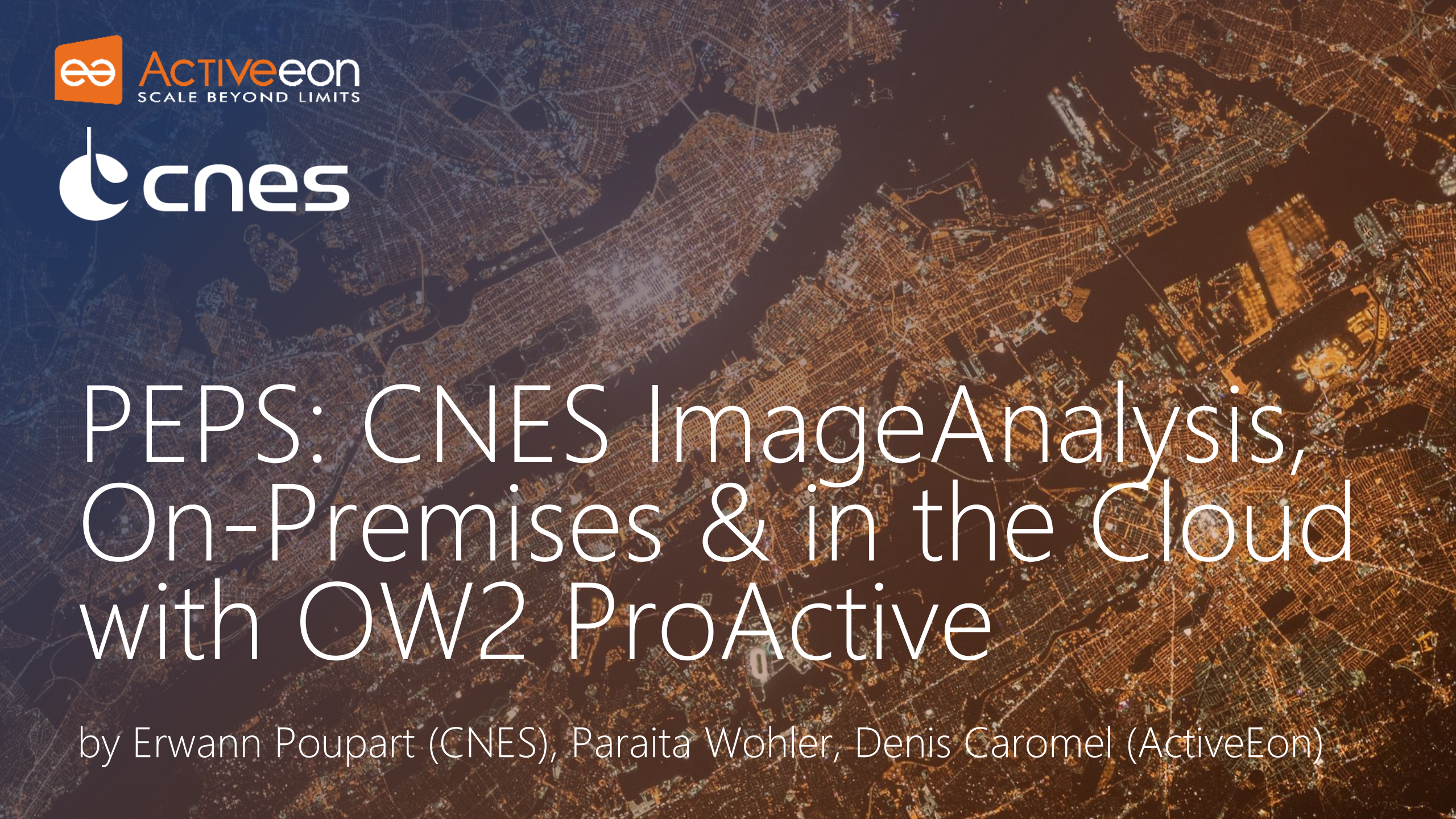


PEPS: CNES ImageAnalysis, On-Premises & in the Cloud with OW2 ProActive

by Erwann Poupart (CNES), Paraita Wohler, Denis Caromel (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

Global Locations

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

ProActive Solution

Scheduling
Orchestration
Meta-scheduling
Resource Allocation

On-premises and on All Clouds
Open Source

Partnerships

amazon web services Partner Network
Google Cloud Platform
redhat. BUSINESS PARTNER
Microsoft Partner Network
docker

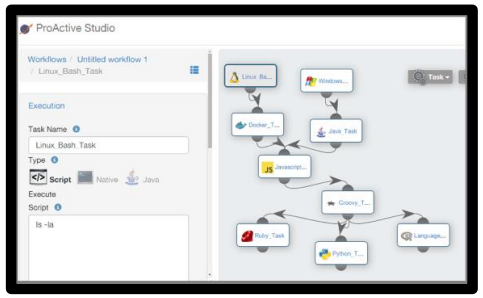
Next Generation Scheduler/Orchestration

ACTIVEEON SCALE BEYOND LIMITS

</> Open REST API

Processing and Automation Workflows

- Any language: python, C:\, R, Java
- Secured Data Transfers
- Meta-scheduler: Control-M, UC4, ca technologies
- ETL, ERP, ELT, …: SAS, SAP, talend
- Full integration
- Translator: ca technologies, sU, Control-M

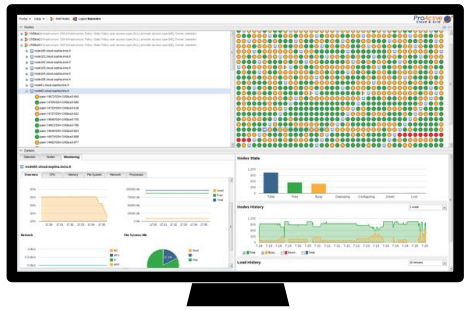


Scheduler and Orchestration

- Priority & Planning
- Parallel Executions
- Error Management
- Multi Users

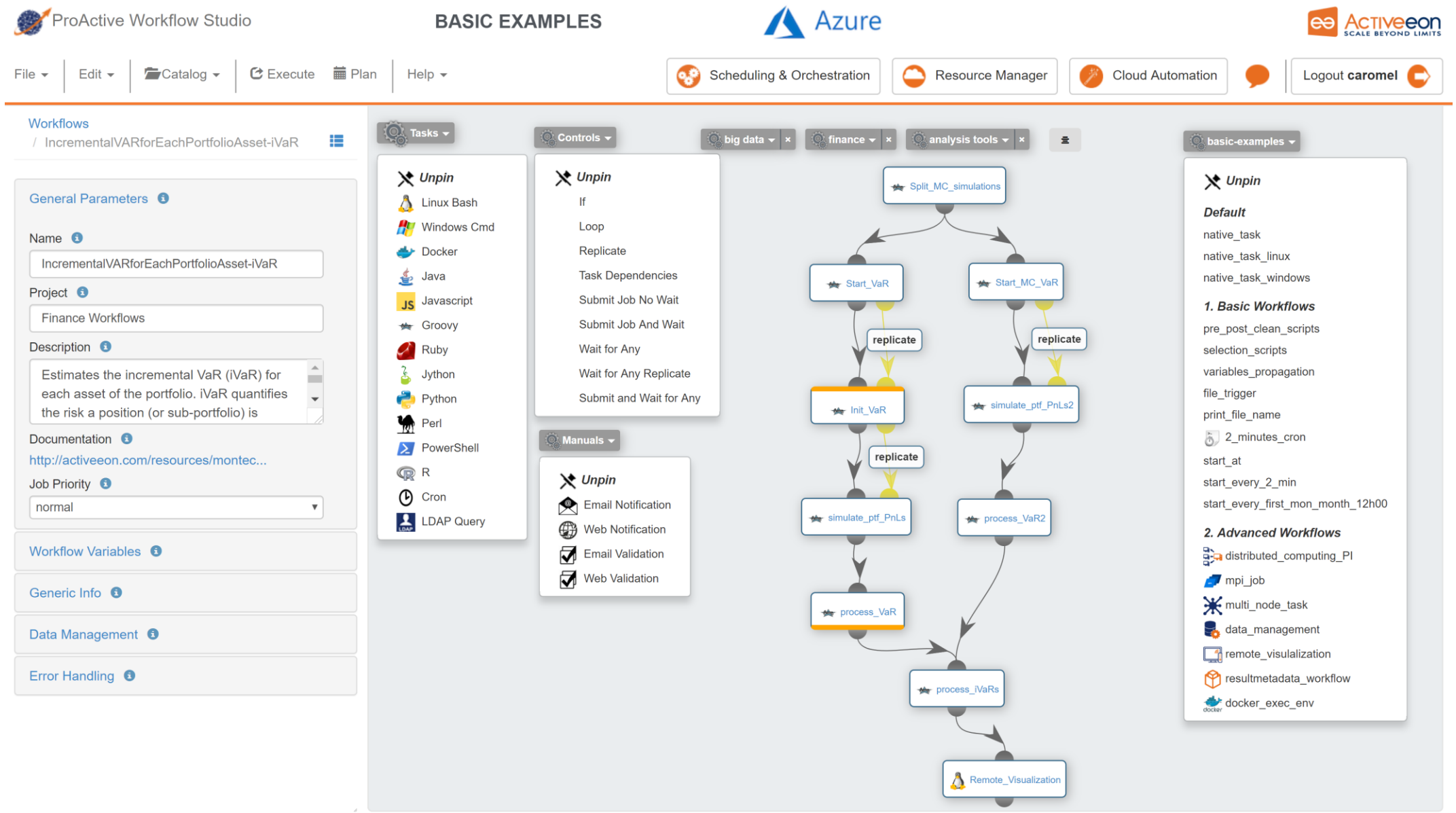
Resource Management and Monitoring

- Multi-platform
- Local Machine
- Network Resource
- Batch Scheduler: Slurm, PBS, SGE, LSF
- Cloud: amazon web services, vmware, openstack, IBM Bluemix, docker, Microsoft Azure, Google

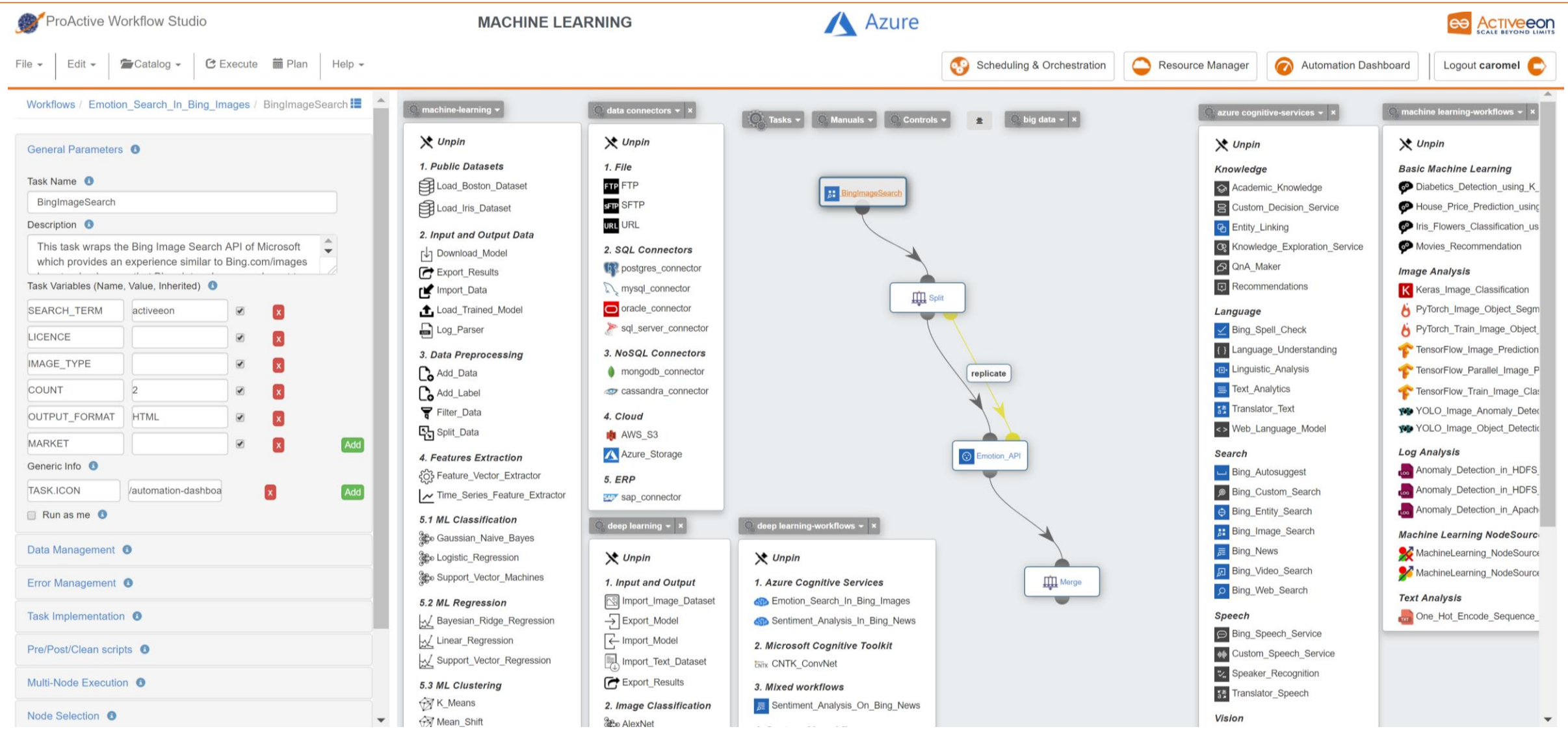


ACTIVEEON SCALE BEYOND LIMITS

Open Workflow Studio

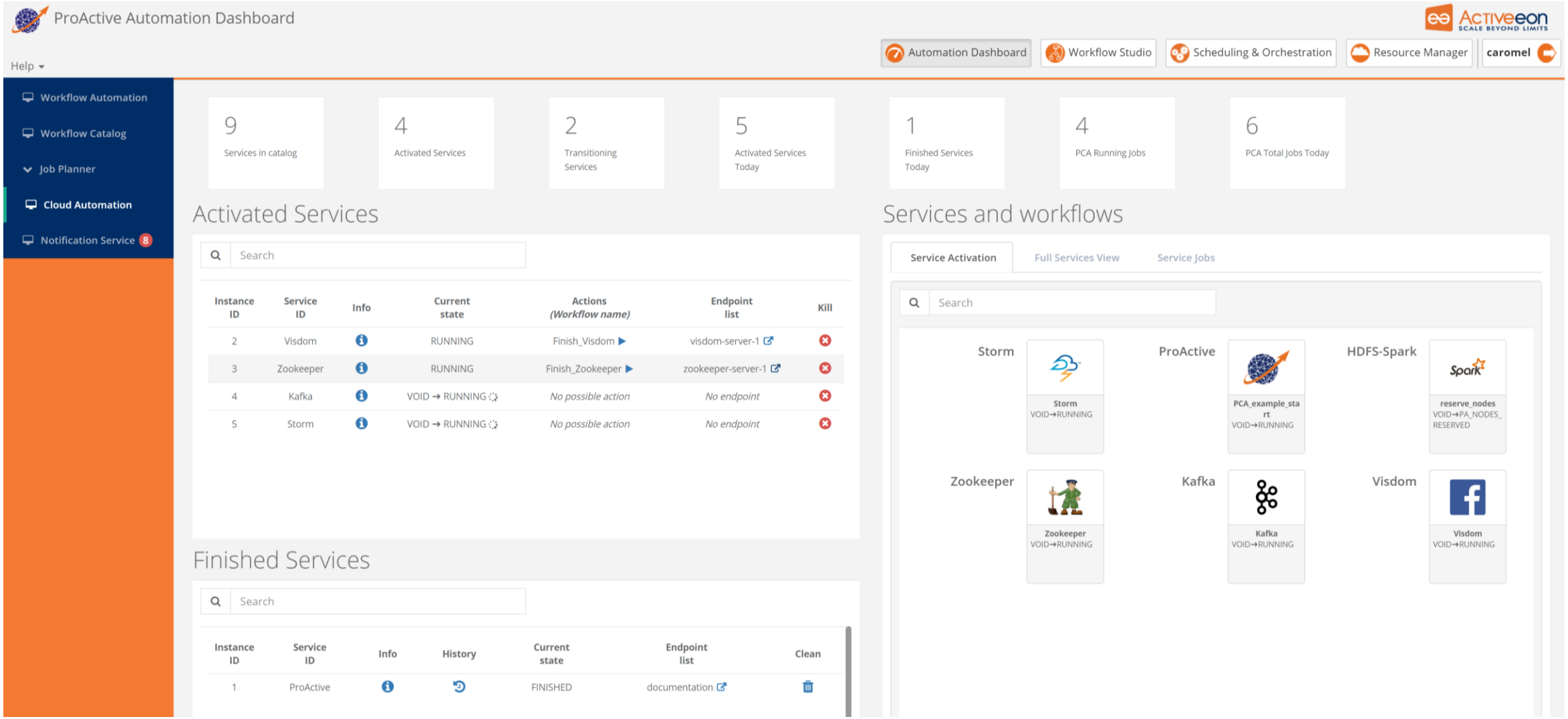


Machine Learning Open Studio



https://www.youtube.com/watch?v=mbrQxCf4lqM

Cloud Automation: On-demand PaaS



On-Demand PaaS Services with full Life-Cycle Management

ProActive Automation Dashboard

ACTIVEEON SCALE BEYOND LIMITS

17	3	0	38	35	3	74
Services in catalog	Activated Services	Transitioning Services	Activated Services Today	Finished Services Today	PCA Running Jobs	PCA Total Jobs Today

Activated Services

Search

Instance ID	Service ID	Info	Current state	Actions *(Workflow name)*	Endpoint list	Kill
36	Elasticsearch		RUNNING	Finish_Elasticsearch	elasticsearch-server-1	
37	Logstash		RUNNING	Finish_Logstash	logstash-server-1	
38	Kibana		RUNNING	Finish_Kibana	kibana-server-1	

Finished Services

Search

Instance ID	Service ID	Info	History	Current state	Endpoint list	Clean
1	PostgreSQL			KILLED	*No endpoint*	
2	Elasticsearch			FINISHED	elasticsearch-server-1	
3	PostgreSQL			KILLED	*No endpoint*	
4	Logstash			FINISHED	logstash-server-1	
5	Kibana			FINISHED	kibana-server-1	
6	MySQL			FINISHED	mysql-server-1	

Services and workflows

Service Activation | Full Services View | Service Jobs

Search

Storm — Storm: VOID→RUNNING

Kibana — Kibana: STOPPED→RUNNING, VOID→RUNNING

ProActive — PCA_example_start: VOID→RUNNING

H2O — H2O: STOPPED→RUNNING, VOID→RUNNING

HDFS-Spark — reserve_nodes: VOID→PA_NODES_RESERVED

PostgreSQL — PostgreSQL: STOPPED→RUNNING, VOID→RUNNING

Clearwater — Clearwater: VOID→RUNNING

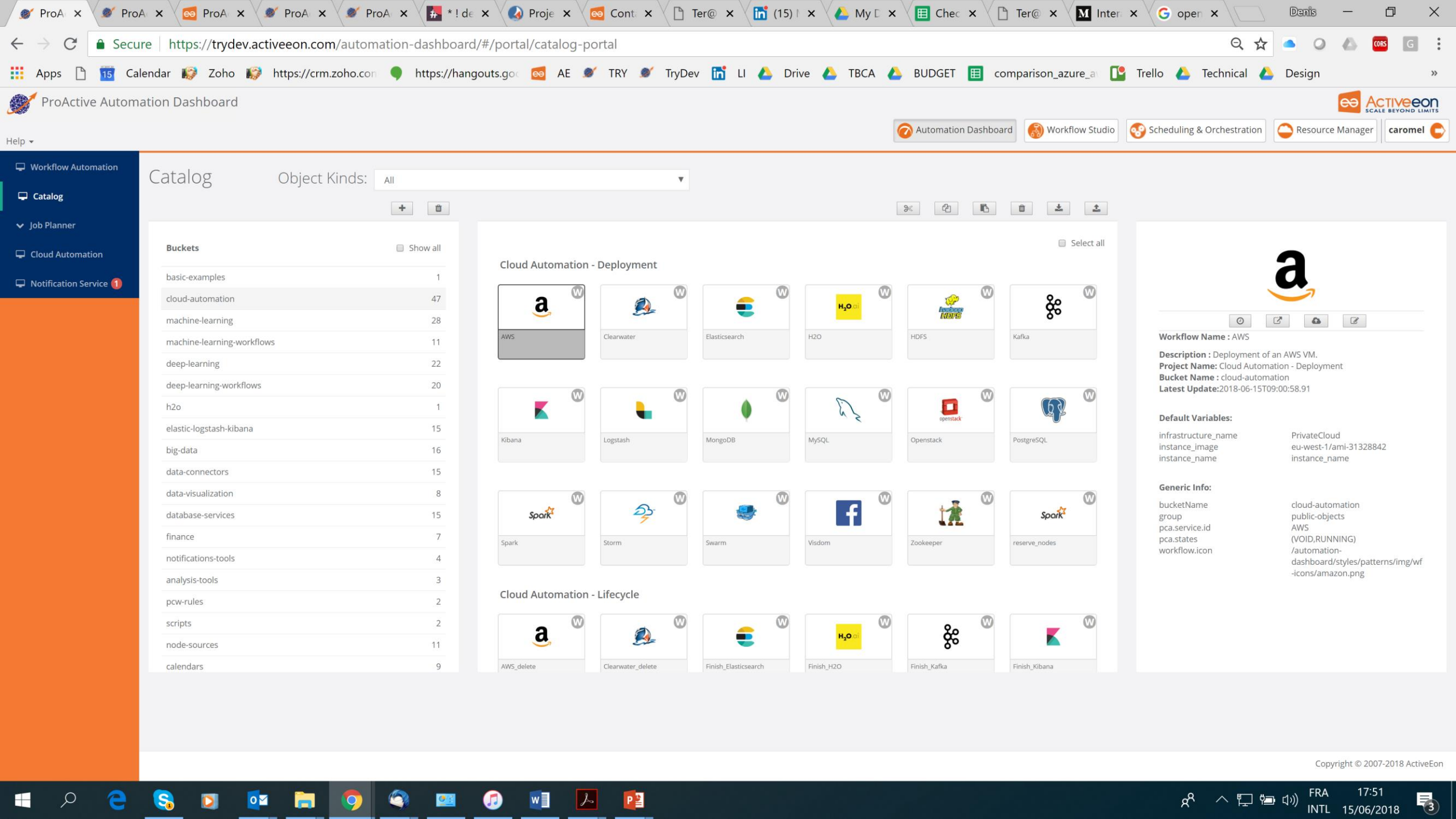
Kafka — Kafka: VOID→RUNNING

Openstack — Openstack: VOID→RUNNING

Logstash

Elasticsearch

PCW-action


ProA
Secure
https://trydev.activeeon.com/automation-dashboard/#/portal/catalog-portal
Apps
Calendar
Zoho
TRY
TryDev
LI
Drive
TBCA
BUDGET
Trello
Technical
Design
ProActive Automation Dashboard
ACTIVEEON
SCALE BEYOND LIMITS
Help
Automation Dashboard
Workflow Studio
Scheduling & Orchestration
Resource Manager
caromel
Workflow Automation
Catalog
Job Planner
Cloud Automation
Notification Service 1
Catalog
Object Kinds:
All
Buckets
Show all
basic-examples 1
cloud-automation 47
machine-learning 28
machine-learning-workflows 11
deep-learning 22
deep-learning-workflows 20
h2o 1
elastic-logstash-kibana 15
big-data 16
data-connectors 15
data-visualization 8
database-services 15
finance 7
notifications-tools 4
analysis-tools 3
pcw-rules 2
scripts 2
node-sources 11
calendars 9
Select all
Cloud Automation - Deployment
AWS
Clearwater
Elasticsearch
H2O
HDFS
Kafka
Kibana
Logstash
MongoDB
MySQL
Openstack
PostgreSQL
Spark
Storm
Swarm
Visdom
Zookeeper
reserve_nodes
Cloud Automation - Lifecycle
AWS_delete
Clearwater_delete
Finish_Elasticsearch
Finish_H2O
Finish_Kafka
Finish_Kibana
Workflow Name : AWS
Description : Deployment of an AWS VM.
Project Name: Cloud Automation - Deployment
Bucket Name : cloud-automation
Latest Update:2018-06-15T09:00:58.91
Default Variables:
infrastructure_name PrivateCloud
instance_image eu-west-1/ami-31328842
instance_name instance_name
Generic Info:
bucketName cloud-automation
group public-objects
pca.service.id AWS
pca.states (VOID,RUNNING)
workflow.icon /automation-dashboard/styles/patterns/img/wf-icons/amazon.png
Copyright © 2007-2018 ActiveEon
FRA
INTL
17:51
15/06/2018

Some Supported Languages and Connectors

ACTIVEEON
SCALE BEYOND LIMITS

Machine Learning & Artificial Intelligence


Cognitive Services


CNTK


Keras


PyTorch


TensoFlow


Scikit-Learn


MXNet


YOLO


H2O


Caffe


Spark MLlib


Pandas


JupyterLab


DeepLearning G4J


DLib


BigDL

Big Data


Zookeeper


Kafka


Azure Databricks


Spark


Hadoop HDFS


Hadoop


Swarm


Storm


Clearwater


Twitter

Specialized


Cuda


Open/CL


FPGA

Visualization


Visdom


Grafana


Kibana

Data Connectors


URL


FTP


SFTP


PostgreSQL


Greenplum


MySQL


Oracle


SQL Server

MongoDB

Cassandra

Azure Storage

Azure Data Lake

AWS-S3

SAP

ElasticSearch

Logstach

Languages and Predefined Tasks

Linux Bash

Cmd

Java

Scala

Javascript

Groovy

Ruby

Jython

Python

Perle

PHP

PowerShell

C++/C#

R

Cron

LDAP

Infrastructure

Linux

Windows

Solaris

AIX AS/400

VMware

Openstack

Clouds

Azure

AWS

GCP

Docker

Kubernetes

OpenShift

HPC Schedulers

Slurm

PBS

LSF

SGE

The Centre national d'études spatiales (CNES)

Data Processing at CNES

De l'espace pour la Terre

Numerical simulation

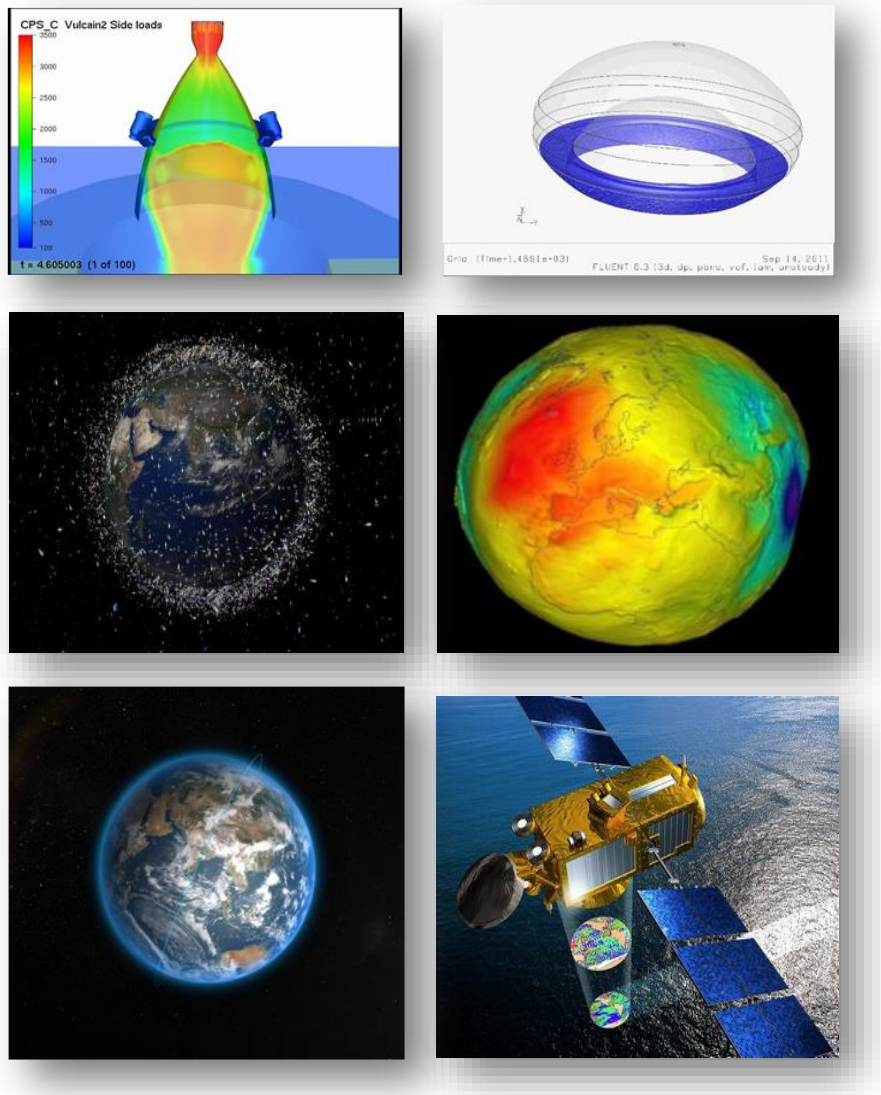
- Research, early phases of missions
- Computation profile mostly HPC centers
- CPU intensive
- Memory bandwidth intensive
- Highly parallel (MPI / OpenMP / CUDA)

On many cores: 8 000 currently

Data processing (HTC)

- Very common to space engineering
- Process satellite raw data into scientific data
- CPU intensive
- Mostly IO centric application
- Coarse grain parallelism (multiple sequential jobs)

On spark, dask, etc.



Microsoft

Activeeon
SCALE BEYOND LIMITS

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.

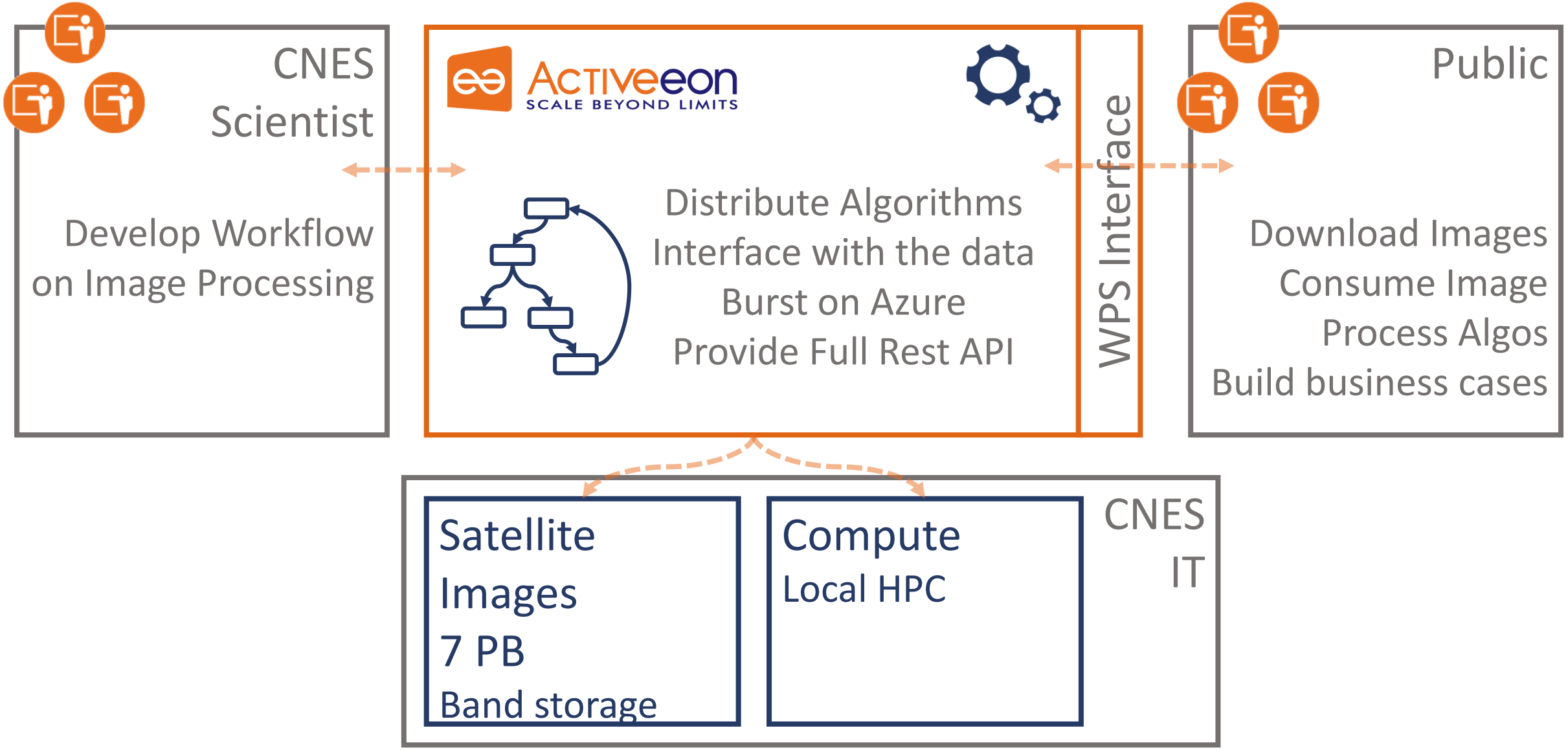
Multi-sensor (radar, optical, etc.), High frequency, long term project.

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

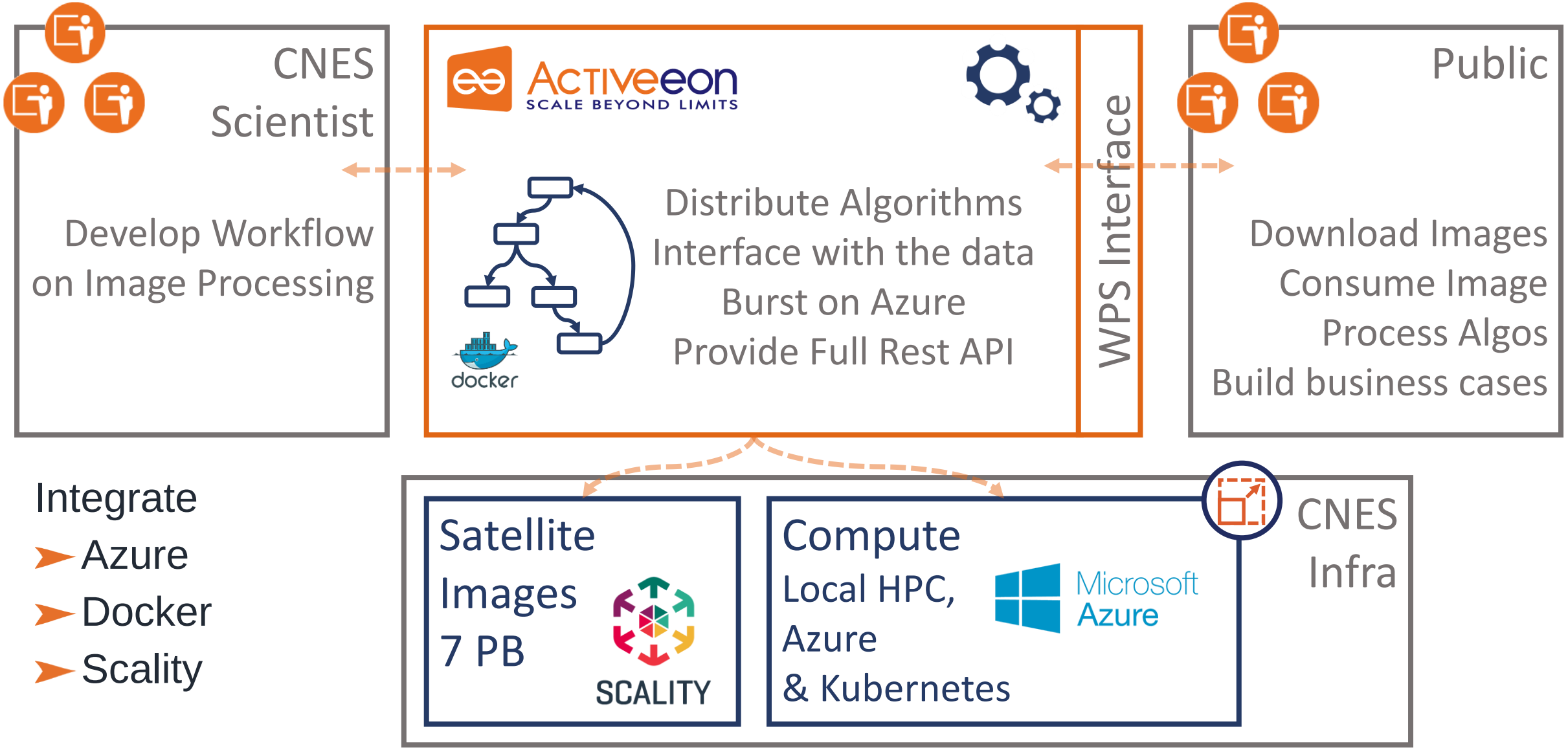
Leverage Cloud Capacity: On-premises to Cloud Burst

Objectives

PEPS Architecture

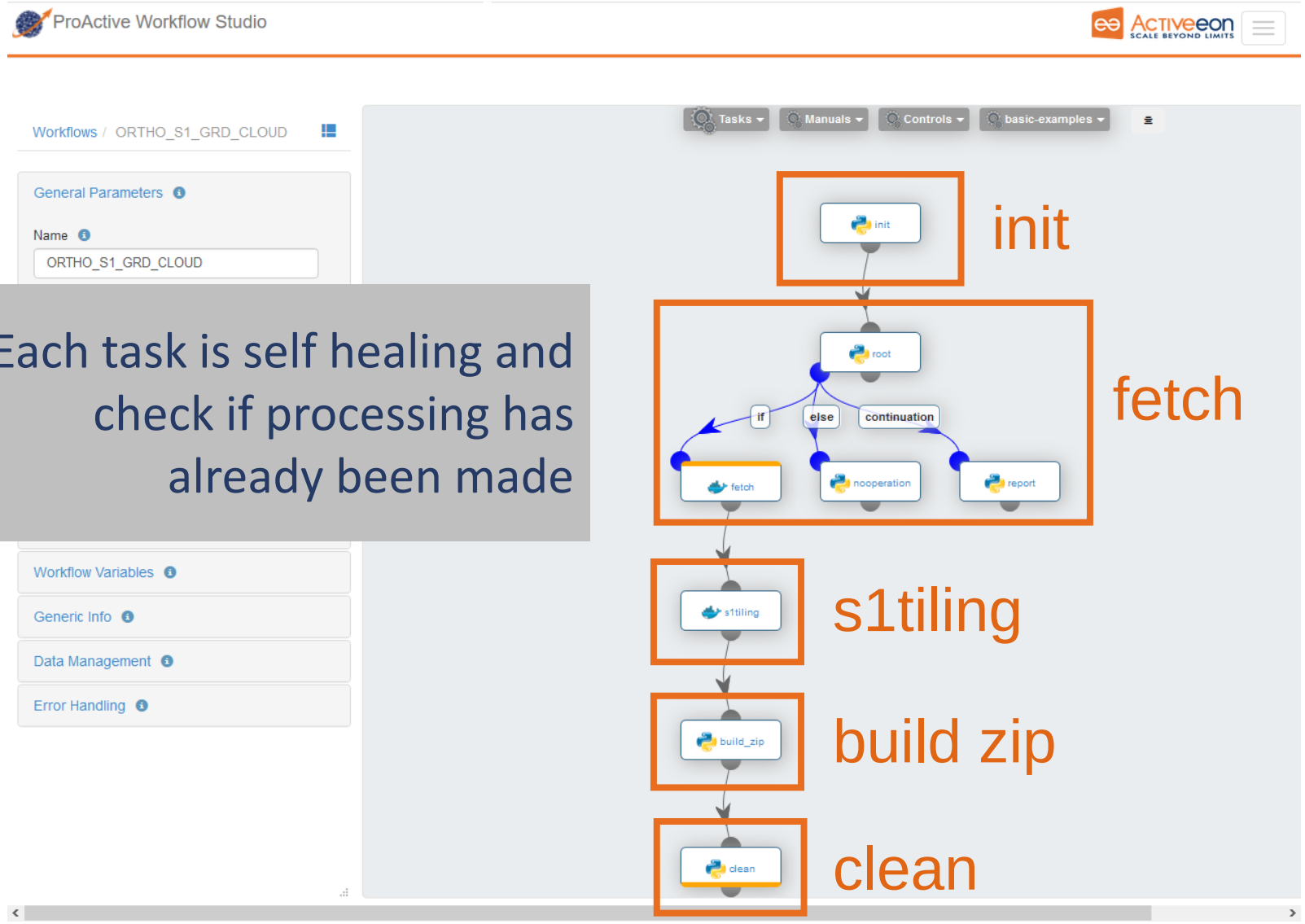


PEPS PoC ecosystem architecture




Microsoft
ACTIVEEON
SCALE BEYOND LIMITS
demo
by cnes

Image Processing Workflow



Each task is self healing and check if processing has already been made

init

Initialize folders and variables

fetch

Fetch data : how and with which parameters according to variables

s1tiling

Image processing over 8 CPU cores

Algorithm on Docker container

Build zip

Compress output

Clean

Demo

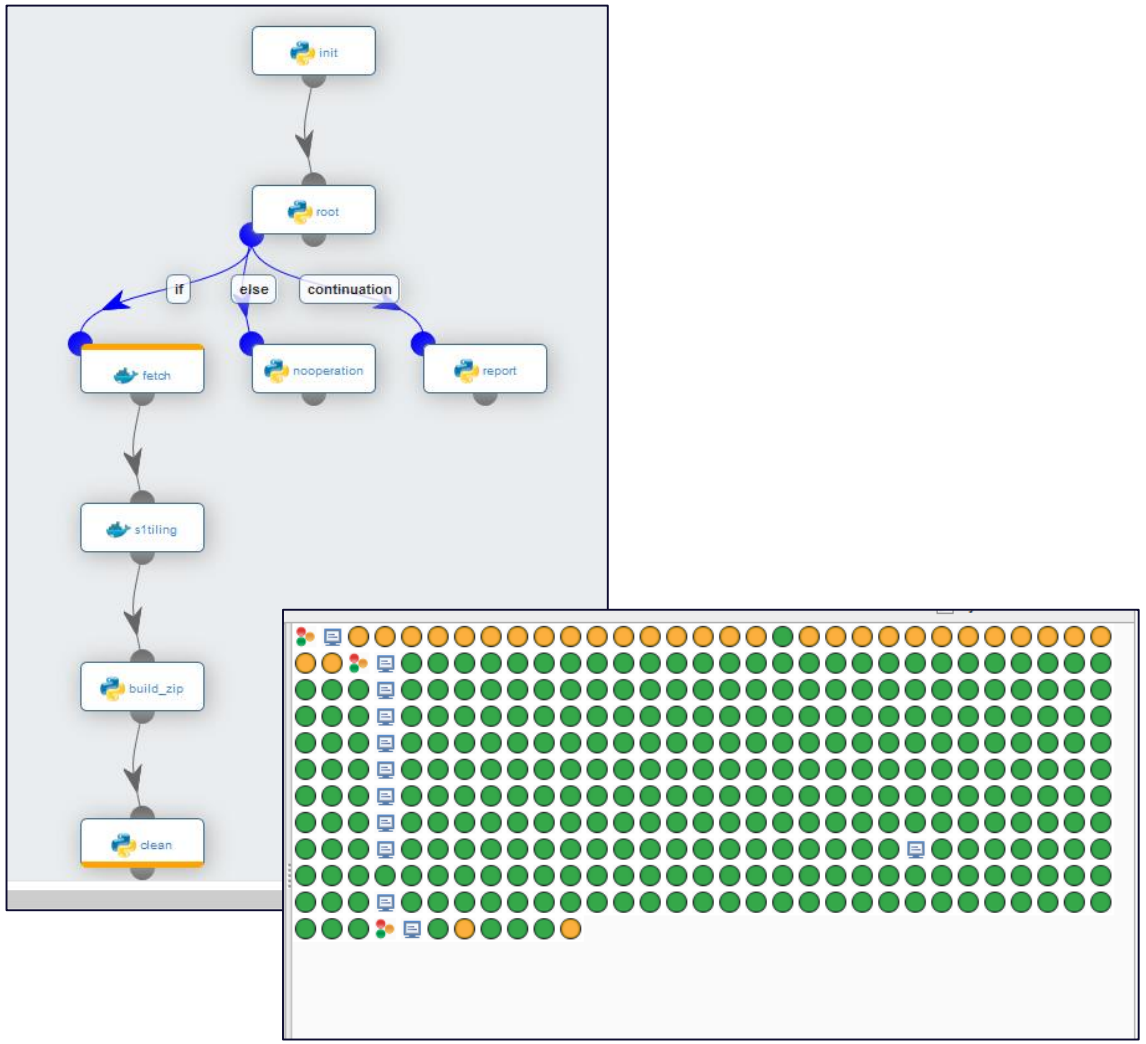
0. Prepare Azure Resources with ScaleSet, 10 VMs with each 32 CPU cores

1. Fetch data from portal

2. Trigger image processing workflow with relevant parameters through WPS

3. Provide data to end user

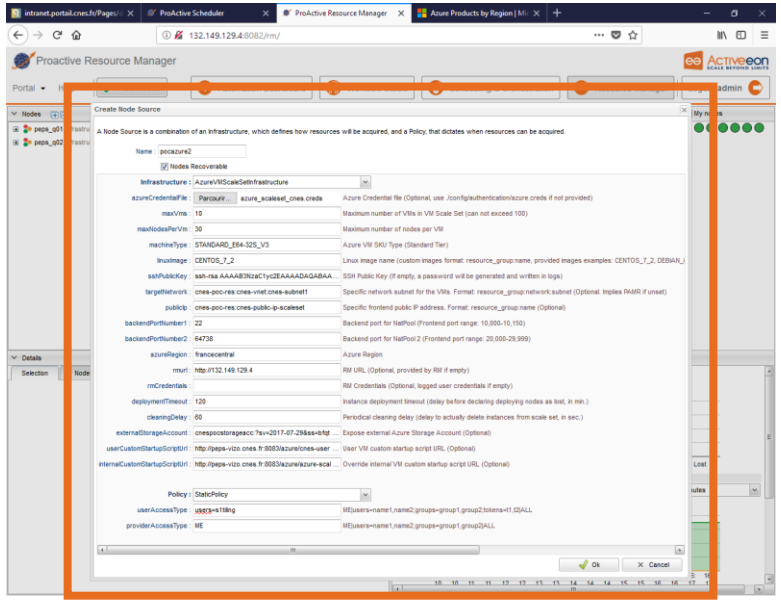
4. Close unnecessary nodes

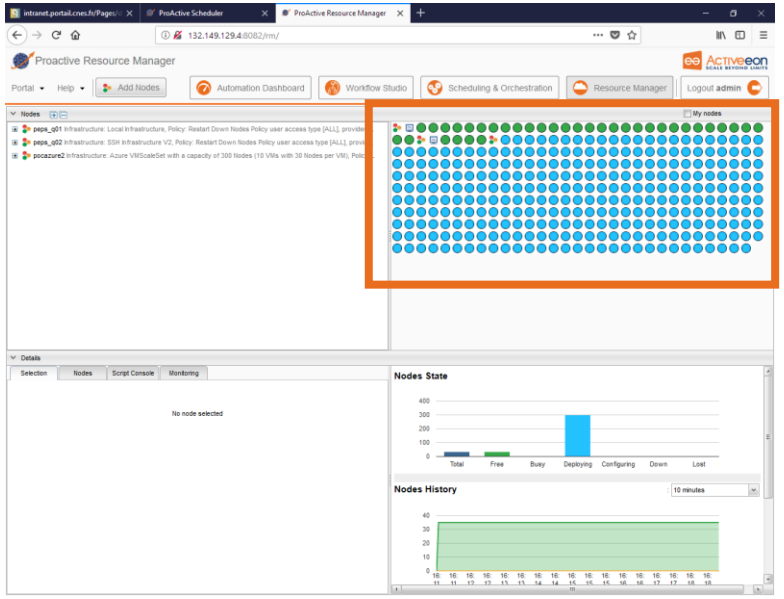


Resource Acquisition

1

Resource acquisition from the RM



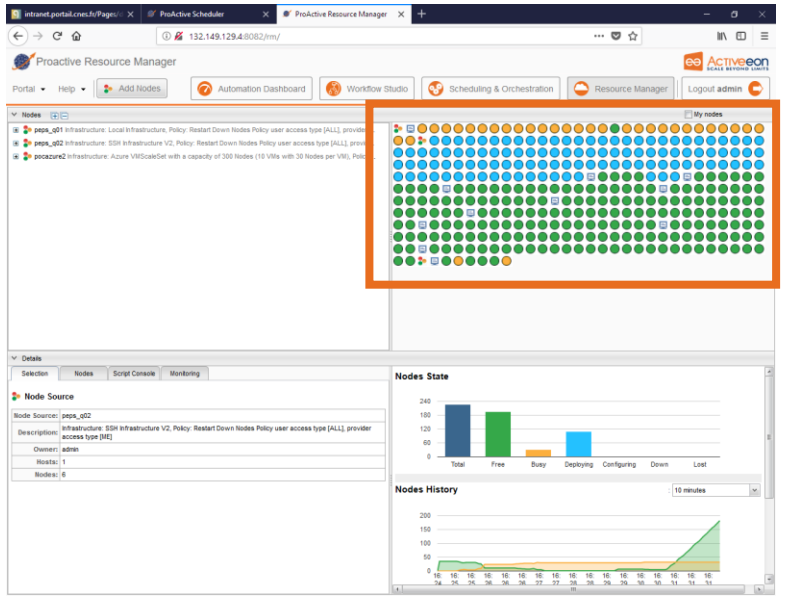


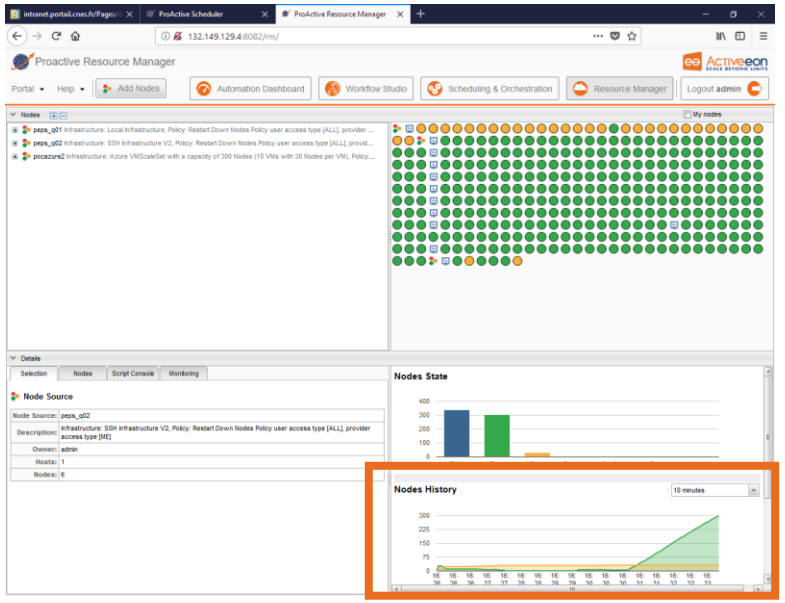
2

Waiting for VMs to start

3

First VMs Connected

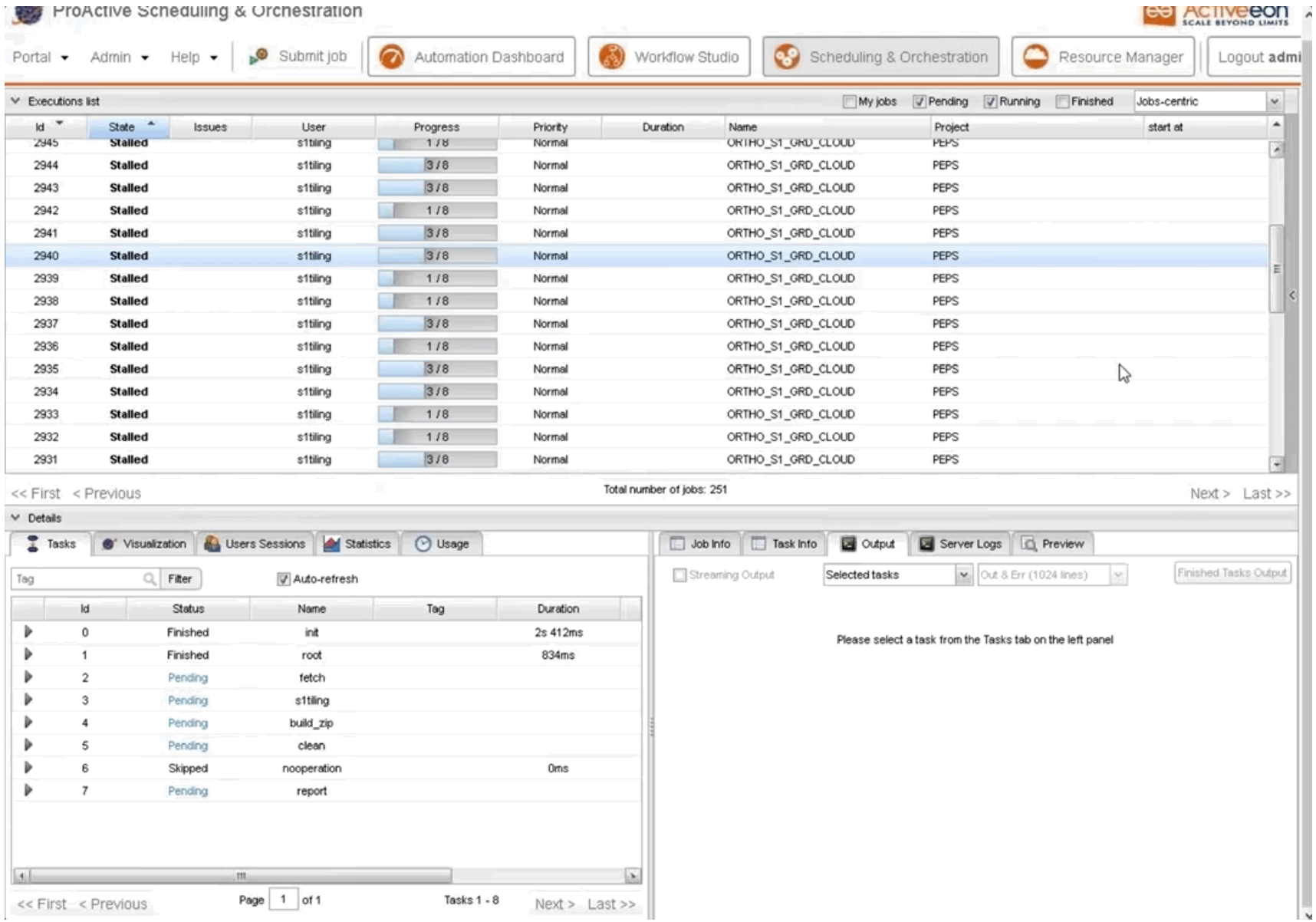




4

Visualize CPU cores available through time

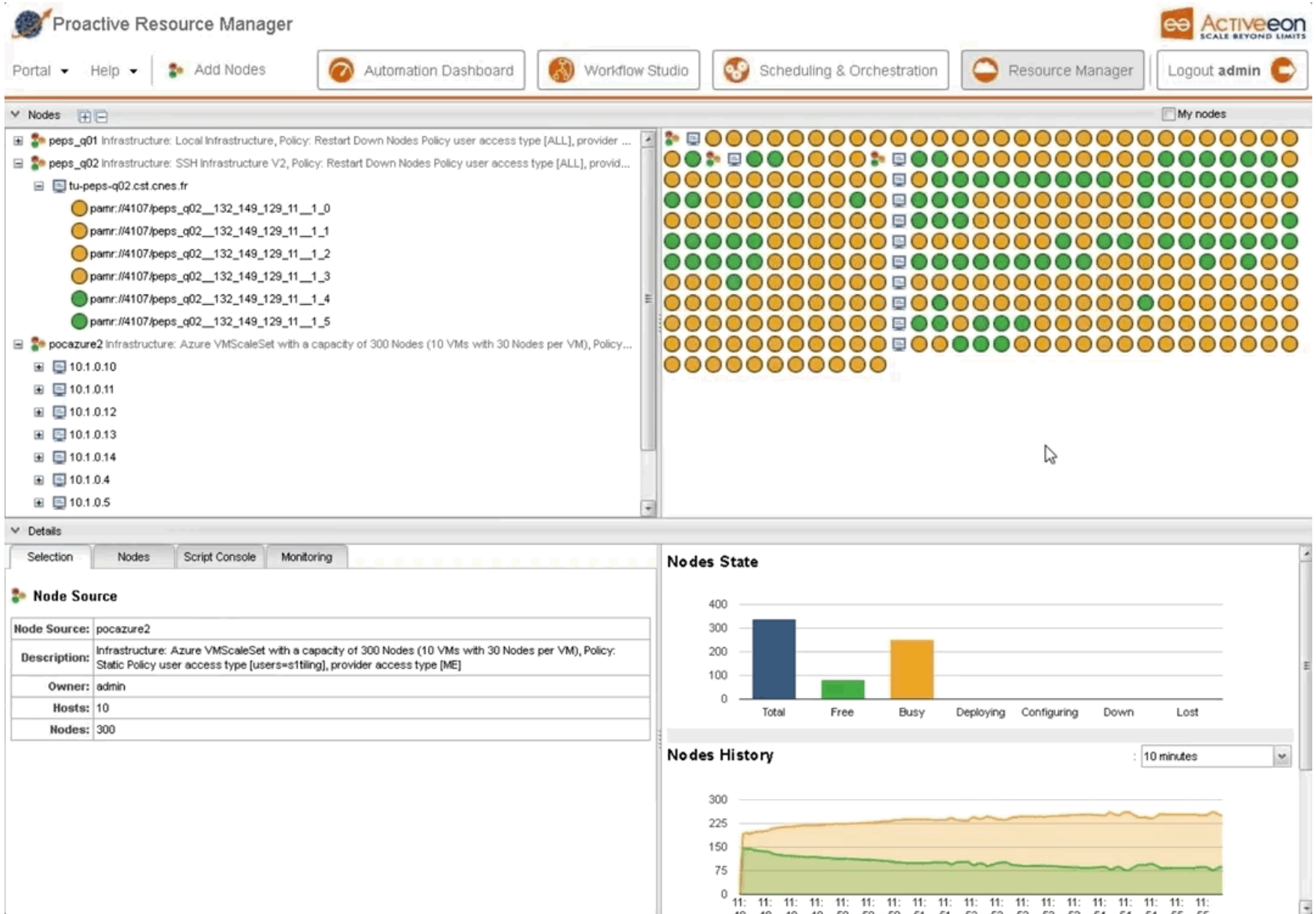
Extensive Parallelization



Distribution over resource pool

- Clear progress visualization
- Parallelization of tasks
- Log retrieval
- Error management

Resource Manager



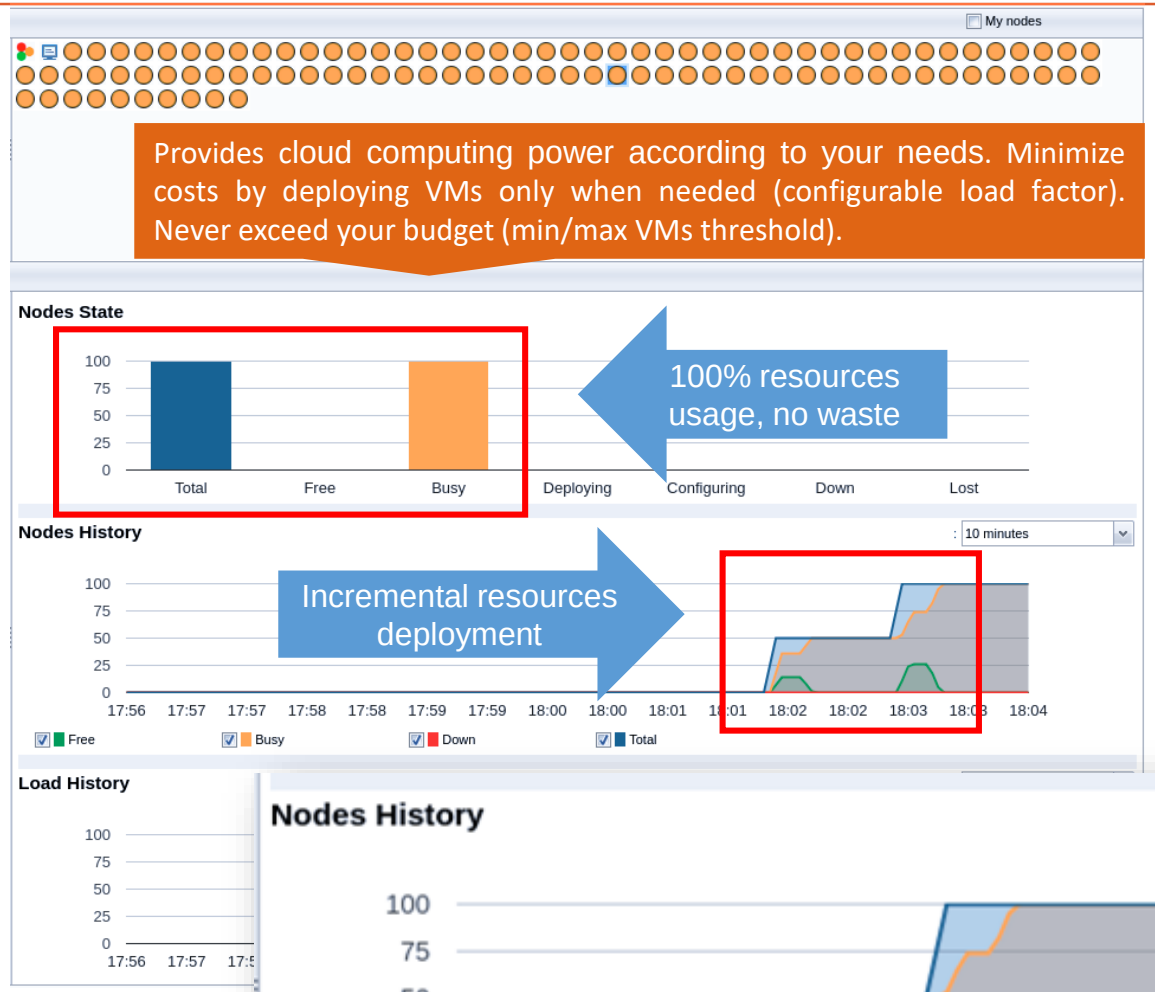
> 20,000 cores

- VM monitoring
- Resource utilization status

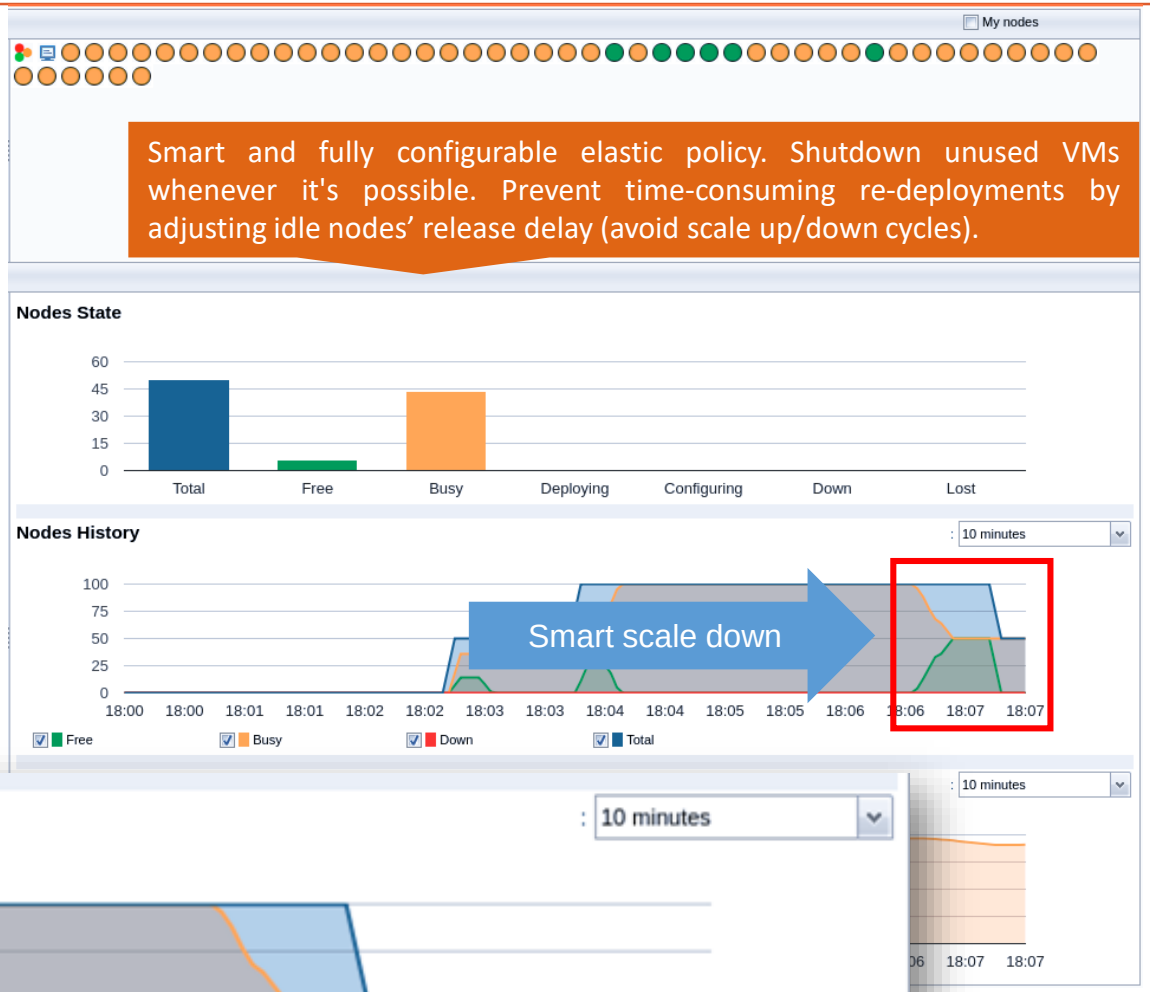
Elasticity: Automatic On-Demand VMs

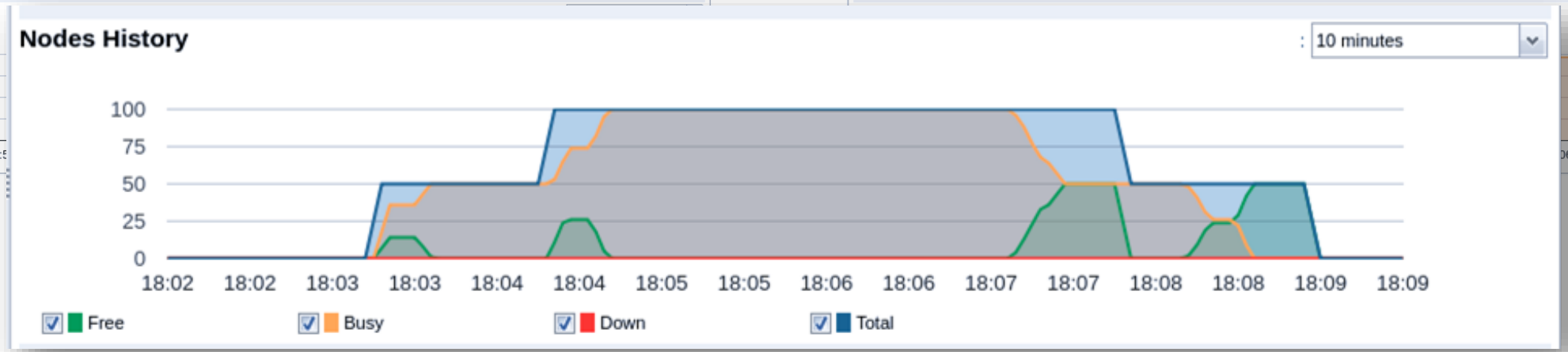
Activeeon
SCALE BEYOND LIMITS

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).



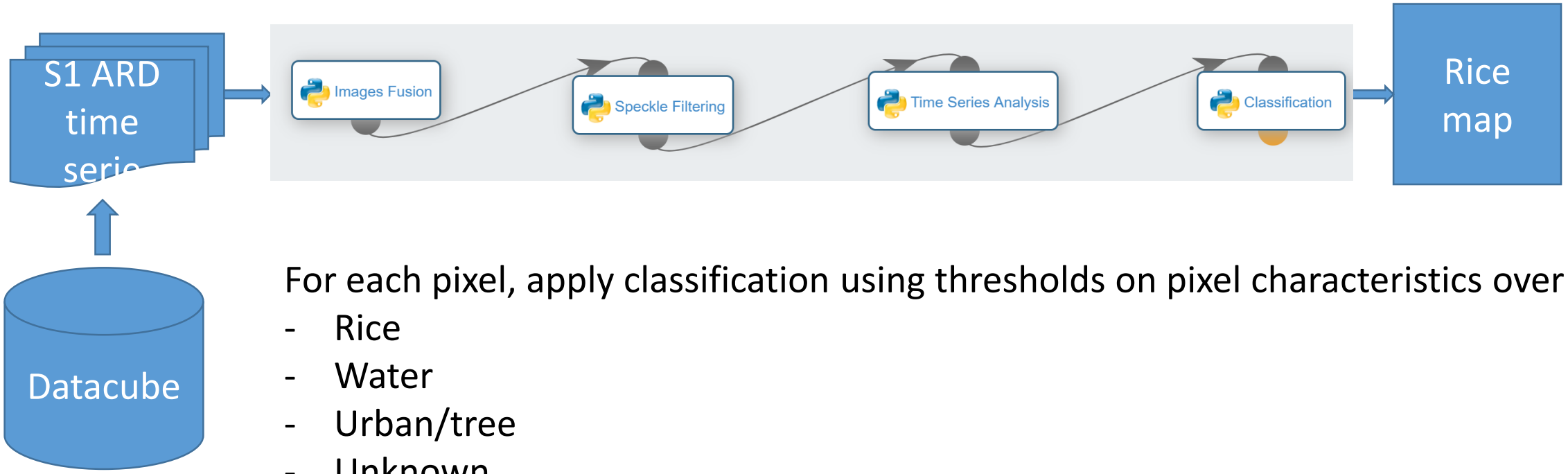
Smart and fully configurable elastic policy. Shutdown unused VMs whenever it's possible. Prevent time-consuming re-deployments by adjusting idle nodes' release delay (avoid scale up/down cycles).





Demonstration of Rice mapping

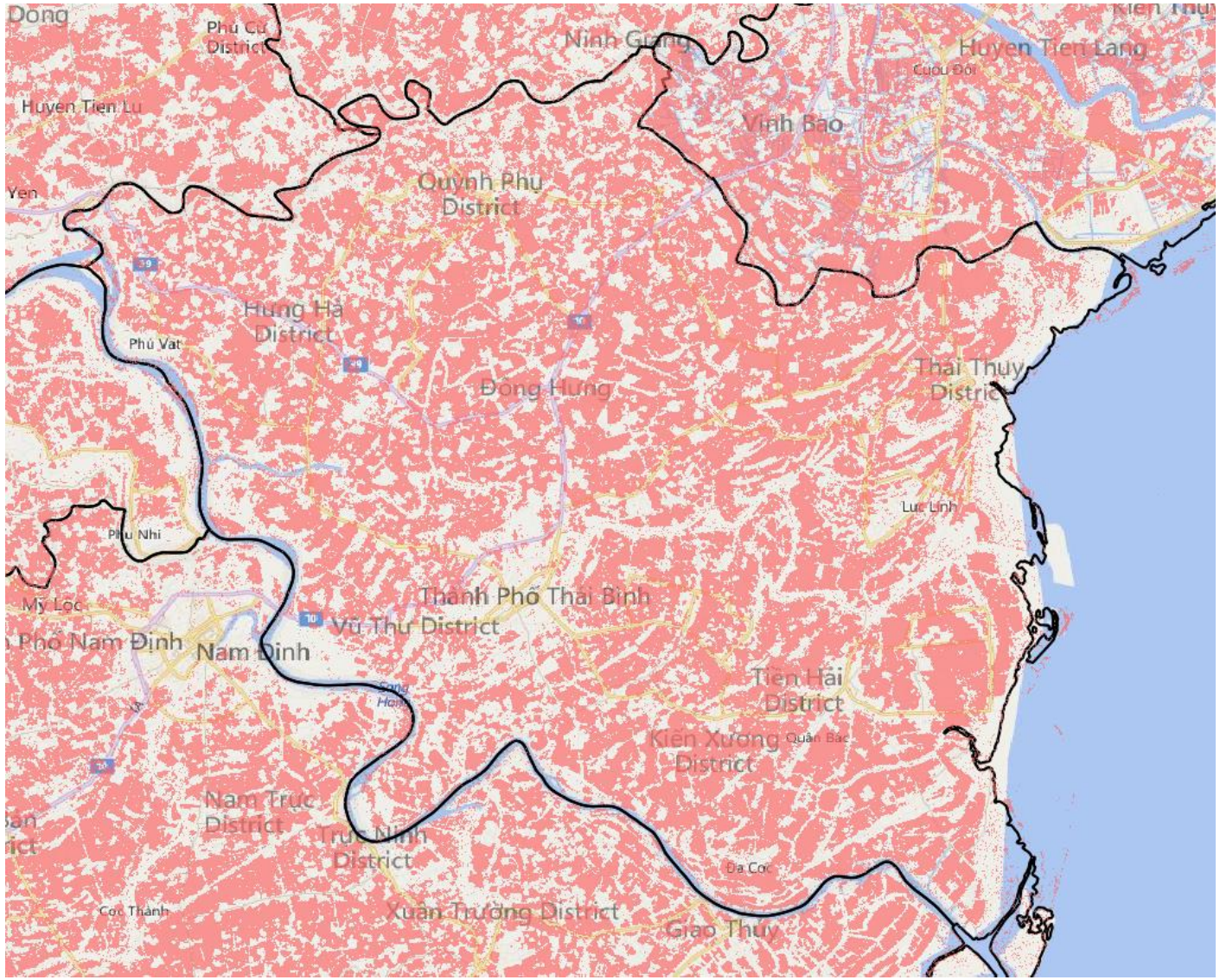
- Objectives: Generate a rice map using Sentinel-1 time series
- Algorithm description:



For each pixel, apply classification using thresholds on pixel characteristics over time:

- Rice
- Water
- Urban/tree
- Unknown

Example of Result: Rice Growth and Harvest Prediction



Winter-Spring Rice Map, Thai Binh province, Thailland

Activeeon
SCALE BEYOND LIMITS

our journey

by cnes

Benefits of ActiveEon

ProActive Integration

- Leverage Docker Images
- Agile development

Activeeon / CNES

Installation on CNES infrastructure

- Automated scripts
- Simplified upgrade

Azure Connection

- Azure expert support
- Security compliance

ActiveEon + Azure:

- Resource agnostic
- Hybrid platform

docker

Activeeon
SCALE BEYOND LIMITS

Microsoft Azure

Benefits of ActiveEon

WPS connector

- Open platform
- Rest API

Connector Integration

- Azure Scale-Set for Advanced resource strategy
- Storage upgrade planned with Scality

To be continued…

ACTIVEEON
SCALE BEYOND LIMITS

SCALITY

Microsoft Azure

Conclusion

Wrap Up

Objective for Sentinel Satellites:

Make Sentinel data available to the greatest number and encourage the development of applications using them (agriculture, maritime field…). 1 petabyte (1015 bytes) in 20 years and now 7 petabytes in 2 years!

Solution

ActiveEon OW2 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

- Optimisation of On-Prem ressources & Clouds
- Faster execution
- Easier to use by end-users
- Cost Reduction with ActiveEon Elastic Provisioning

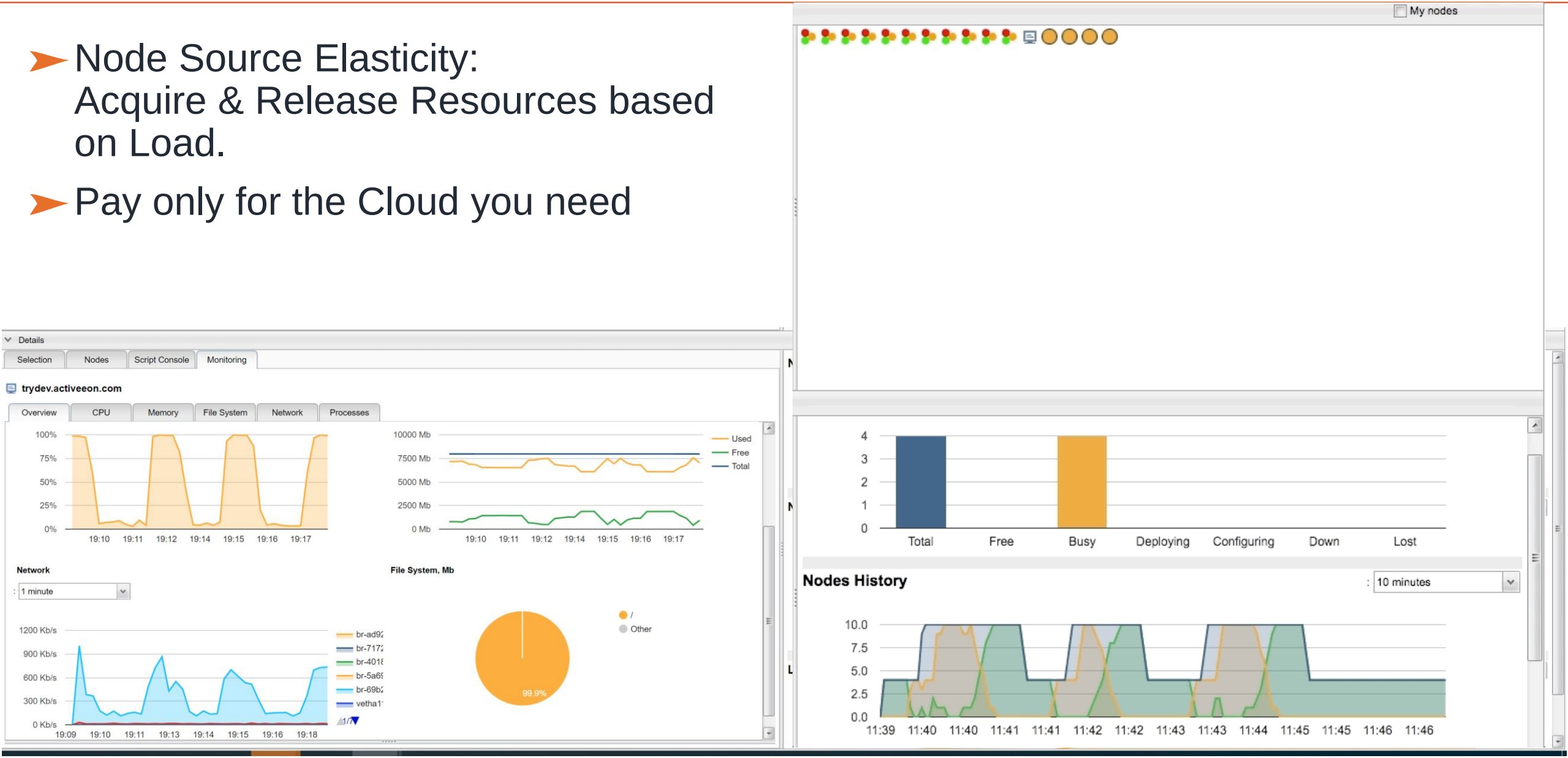

ACTIVEEON
SCALE BEYOND LIMITS

SnapPlanet

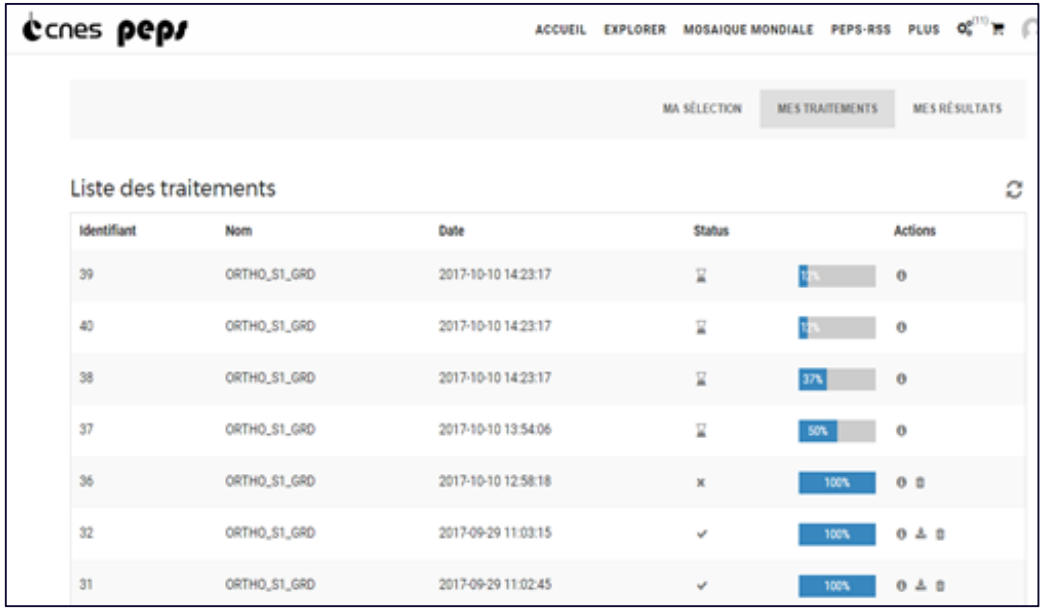
Activeeon
SCALE BEYOND LIMITS

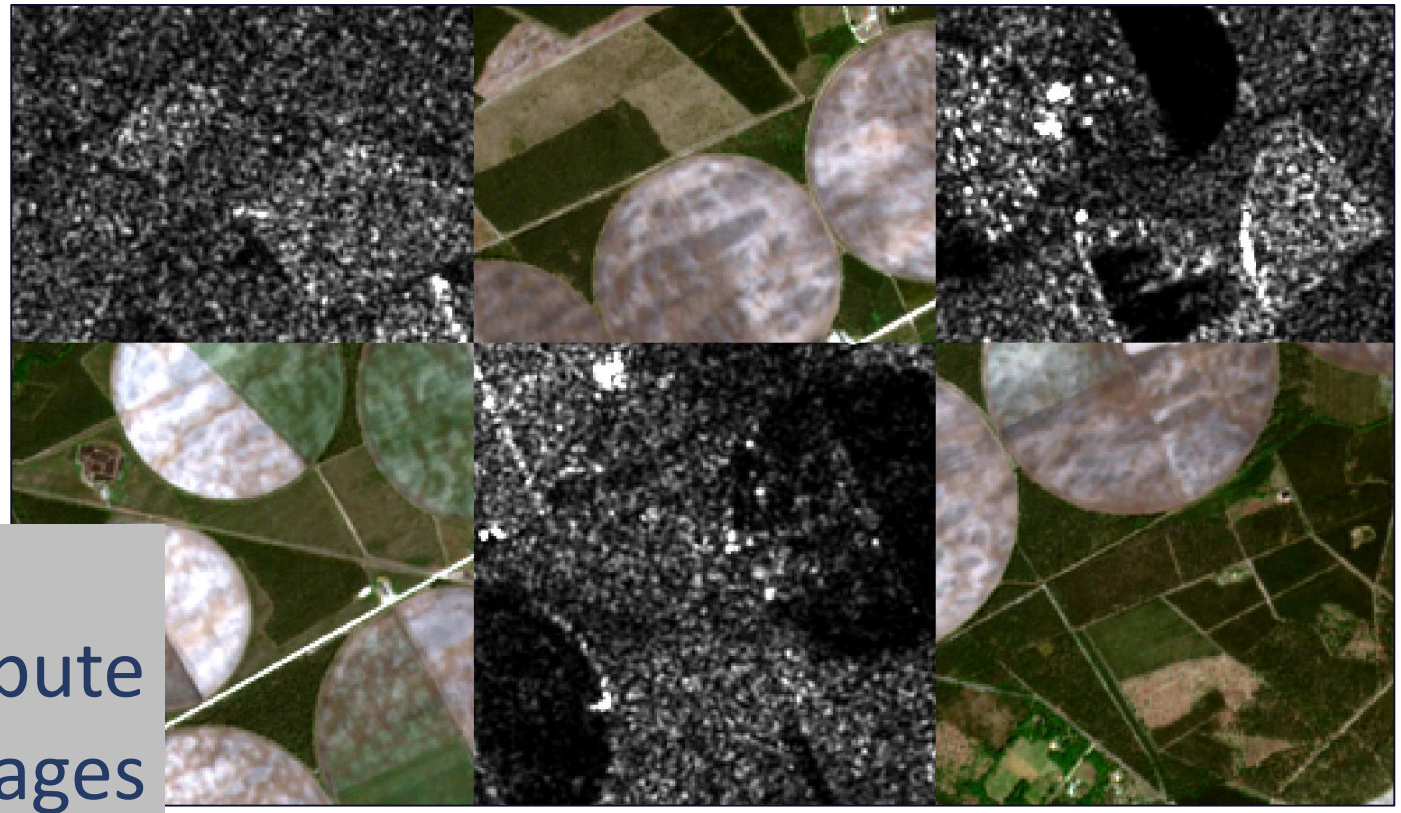
Elasticity: Automatic On-Demand VMs

- Node Source Elasticity:
  Acquire & Release Resources based on Load.
- Pay only for the Cloud you need



Result and Integration

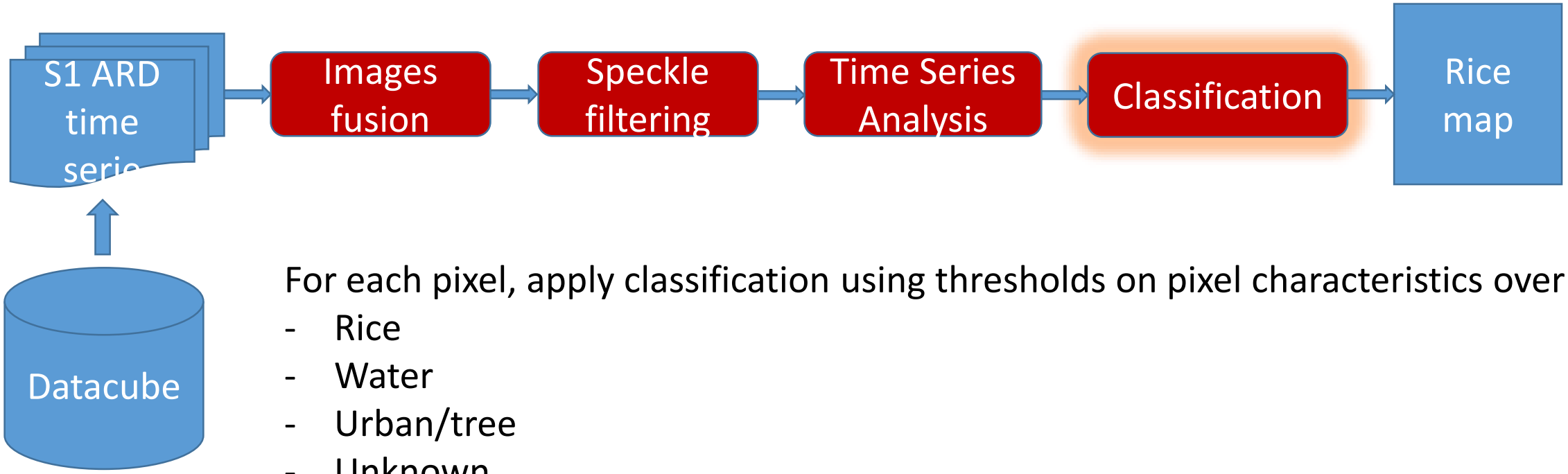




From PEPs Portal access compute power and processed images

Demonstration of Rice mapping

- Objectives: Generate a rice map using Sentinel-1 time series
- Algorithm description:



For each pixel, apply classification using thresholds on pixel characteristics over time:

- Rice
- Water
- Urban/tree
- Unknown

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

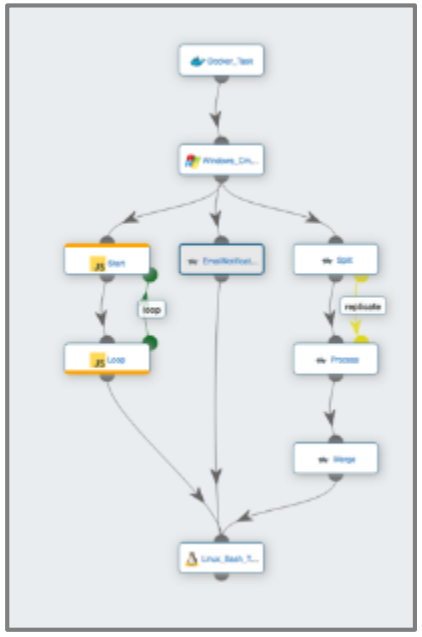
Benefits

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

On-Prem & Hybrid- Multi-cloud Orchestration

ACTIVEEON
SCALE BEYOND LIMITS

Scale automatically - Leverage All Resources

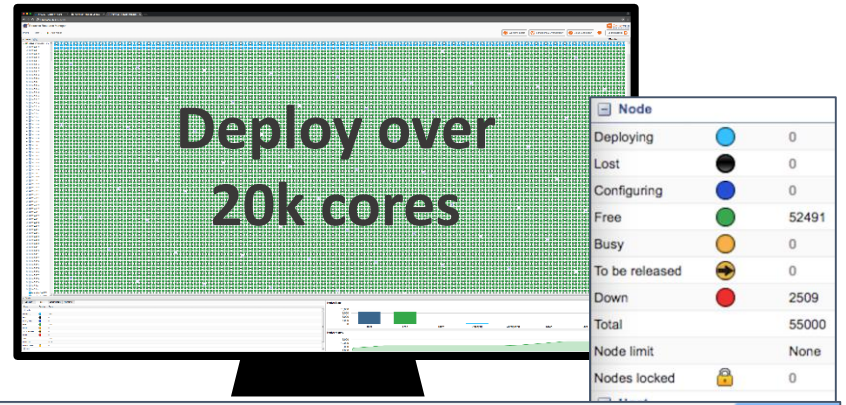


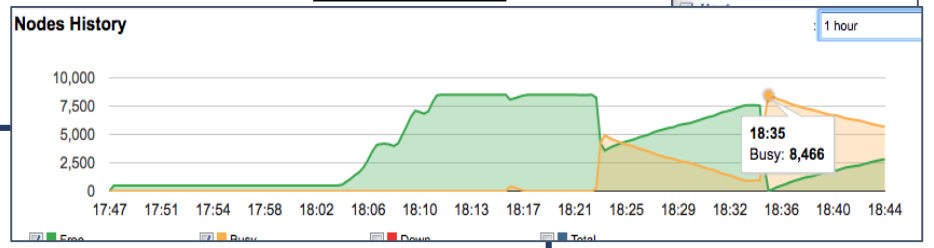
hadoop · kubernetes · docker · Java · Spark · python · SAS · GREENPLUM DATABASE · ANACONDA · kafka

ProActive Workflows & Scheduling

Workflow Scheduler

Resource Manager





Azure Node Sources

VM · VM

AWS Node Sources

Amazon S3 · Autoscale · EC2 · EC2

Private Node Sources

openstack · slurm · PBS · docker · vmware