

Des Multi-Cores aux Clouds avec ProActive Parallel Suite

D. Caromel, et al.

Agenda

1. Background: INRIA, ActiveEon
2. ProActive Open Source:
Programming, Scheduling, Resourcing
3. Cloud Seeding with GPU
4. ProActive PACA GRID:
Cloud Portal with GPUs



Parallelism is Complex:
Hiding Infra Details and Seamless Integration with Multi-Cores and GPU

Forum TER@TEC 2010 – 16 juin



1. Background

- ❑ A joint team, Now about 35 persons
- ❑ 2004: First ProActive User Group
- ❑ 2009, April: ProActive 4.1, Distributed & Parallel:
From Multi-cores to Enterprise GRIDs



OASIS Team Composition (35)

□ Researchers (5):

- D. Caromel (UNSA, Det. INRIA)
- E. Madelaine (INRIA)
- F. Baude (UNSA)
- F. Huet (UNSA)
- L. Henrio (CNRS)

□ PhDs (11):

- Antonio Cansado (INRIA, Conic)
- Brian Amedro (SCS-Agos)
- Cristian Ruz (INRIA, Conicyt)
- Elton Mathias (INRIA-Cordi)
- Imen Filali (SCS-Agos / FP7 SC)
- Marcela Rivera (INRIA, Conicyt)
- Muhammad Khan (STIC-Asia)
- Paul Naoumenko (INRIA/Région)
- Viet Dung Doan (FP6 Bionets)
- Virginie Contes (SOA4ALL)
- Guilherme Pezzi (AGOS, CIFR)

□ + Visitors + Interns



Located in Sophia Antipolis, between
Nice and Cannes,
Visitors Welcome!

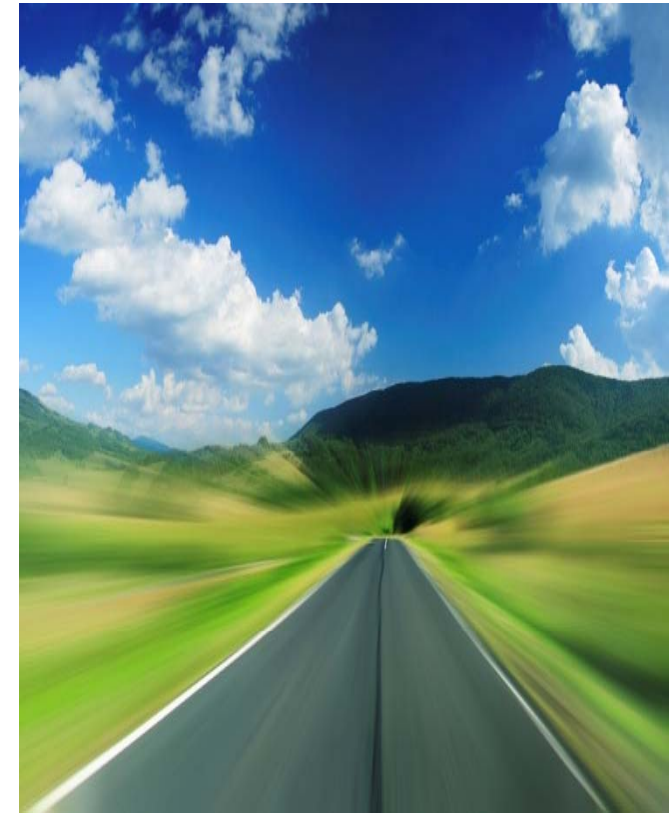
Startup Company Born of INRIA



Some Customers:



Some Partners:

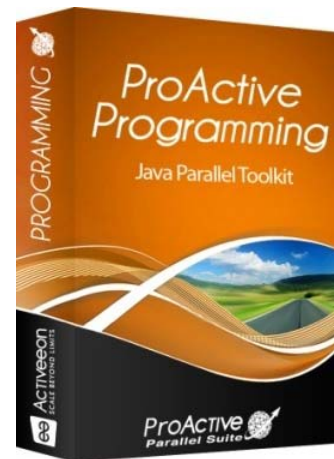
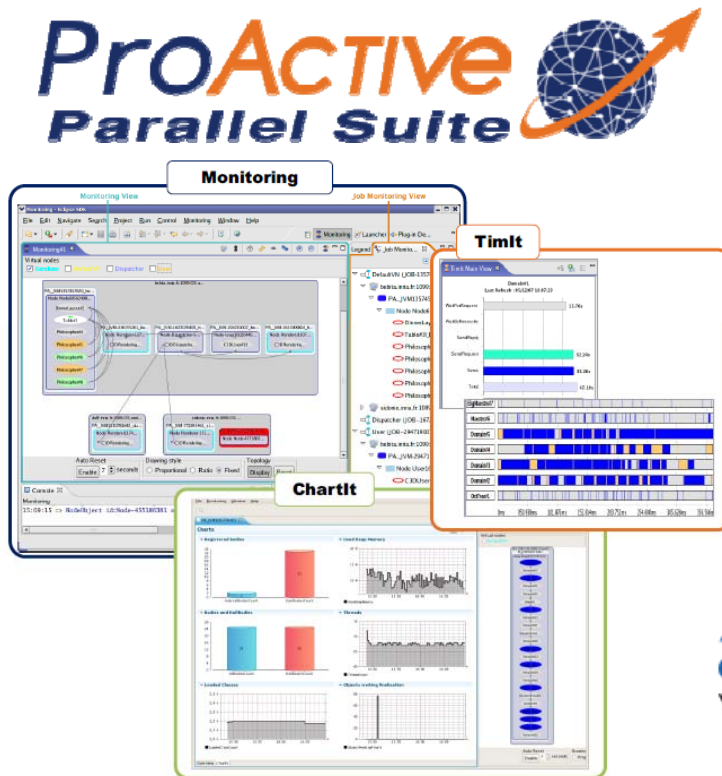


- ❑ Co-developing, Support for [ProActive Parallel Suite](#)
- ❑ Worldwide Customers: Fr, UK, Boston USA

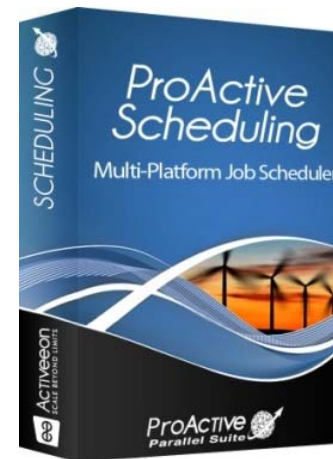


2. ProActive Parallel Suite

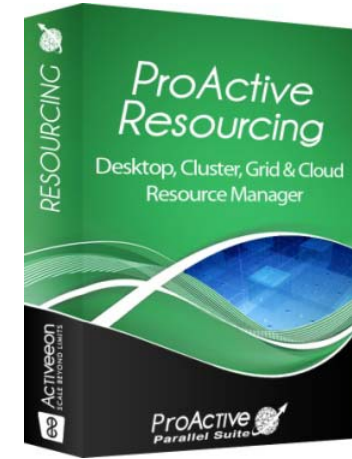
Product: ProActive Parallel Suite



Java Parallel Toolkit



Multi-Platform Job Scheduler



Resource Manager

amaDEUS
Your technology partner

Used in Production Today:
50 Cores → 300 Cores early 2010

Strong Differentiation:

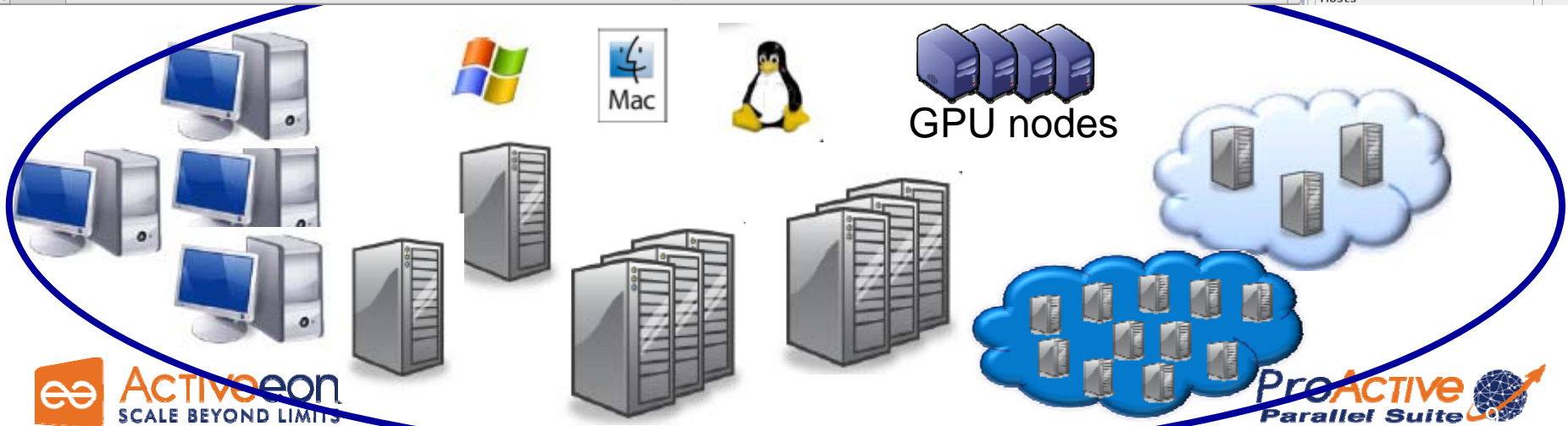
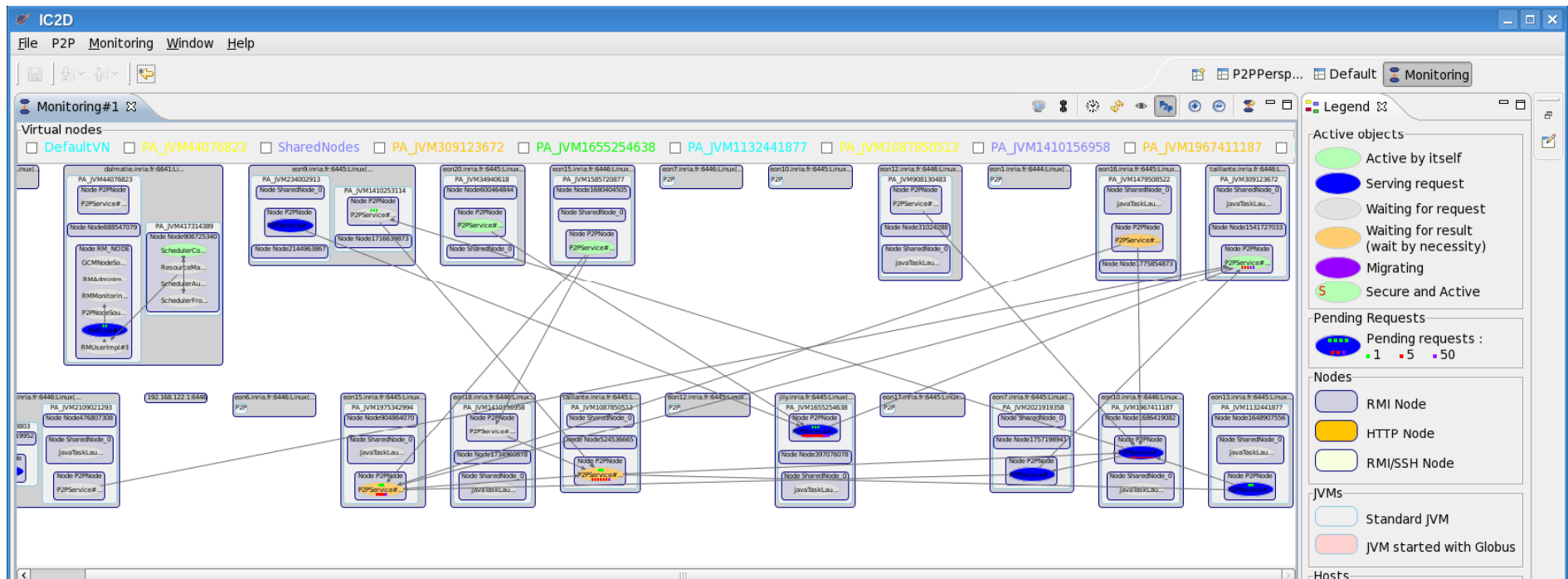
- ❑ Java Parallel Programming + Integration
- ❑ Portability: Linux, Windows, Mac
- ❑ Versatility: Desktops, Cluster, Grid, Clouds

+
+
= Perfect Flexibility

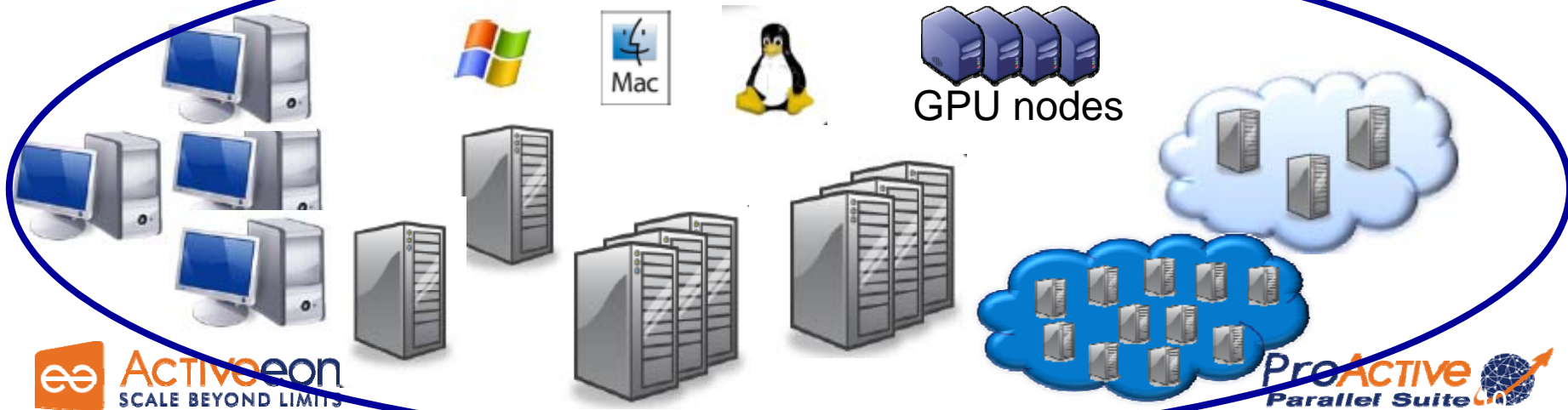
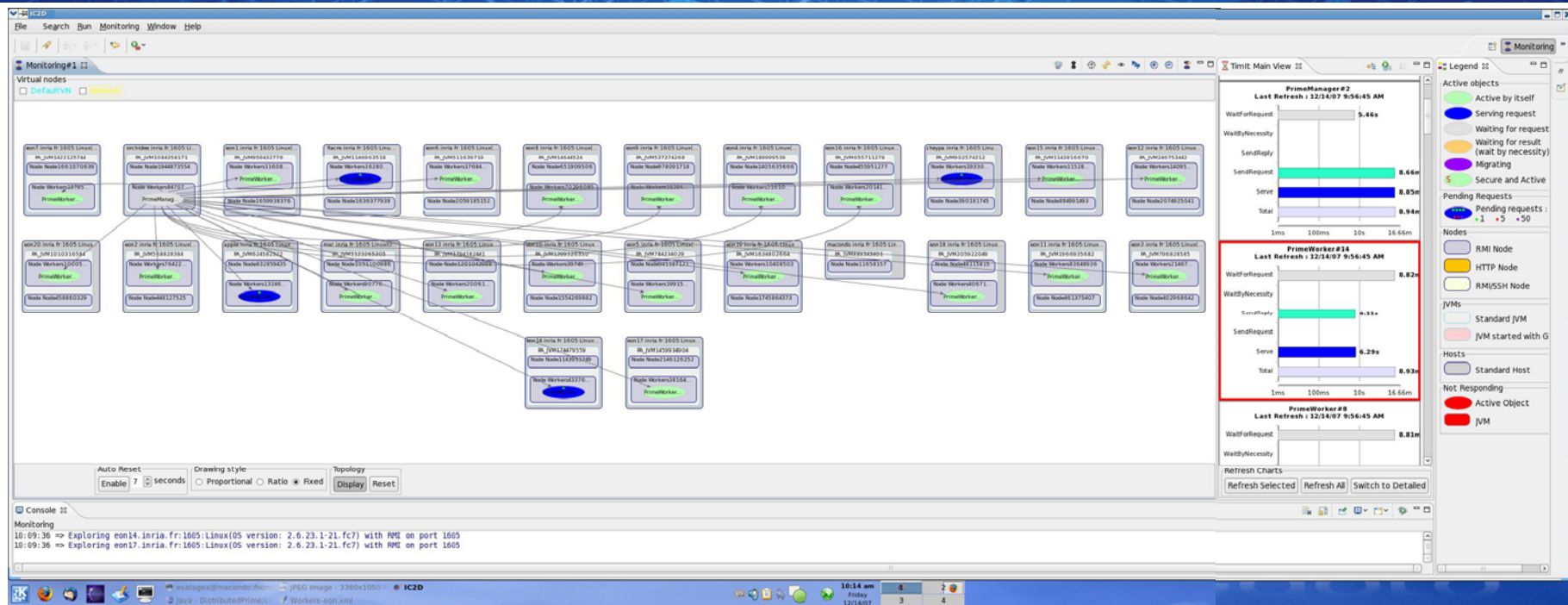


ProActive Programming: Active Objects

ProActive Programming View



ProActive Programming View

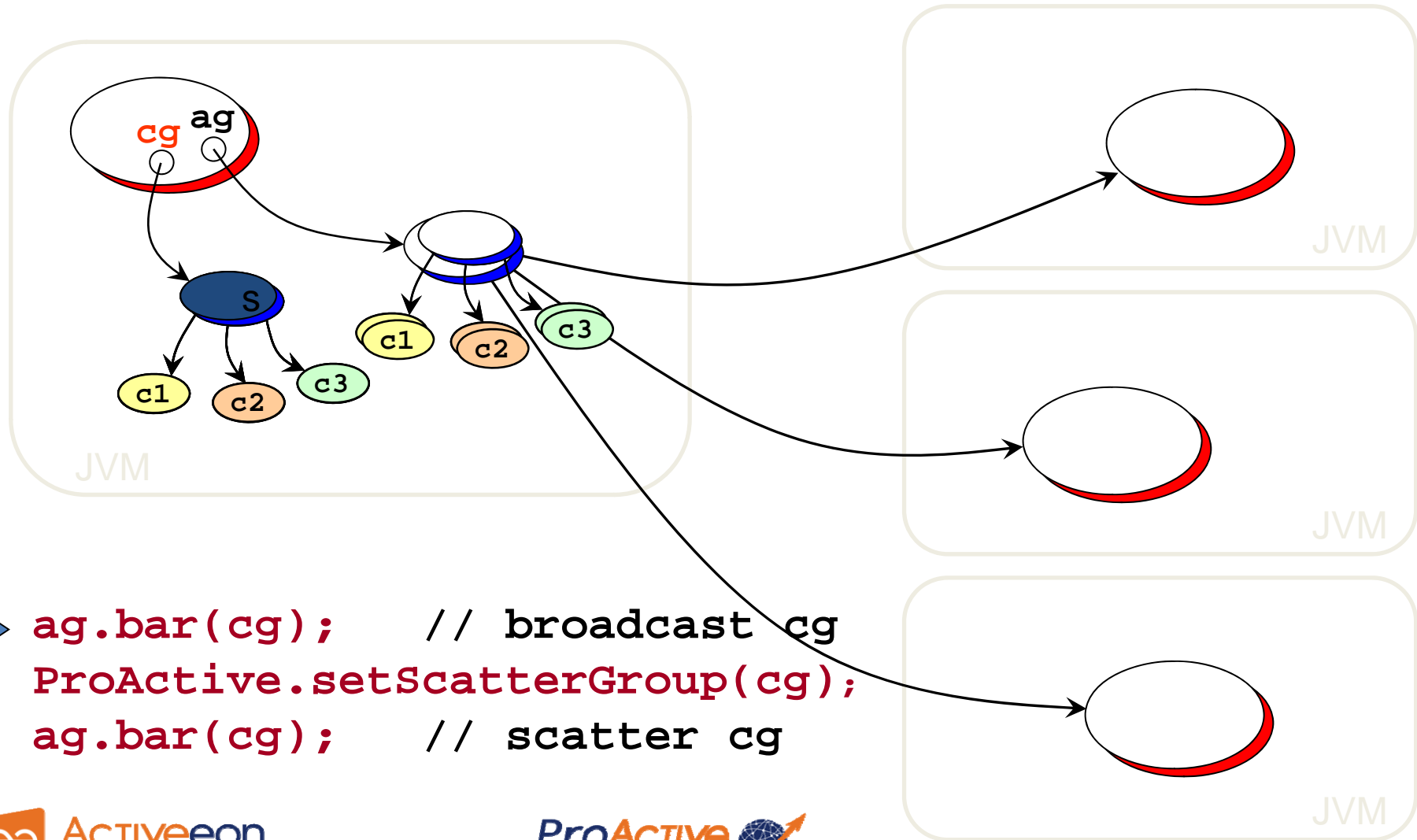


TYPED ASYNCHRONOUS GROUPS

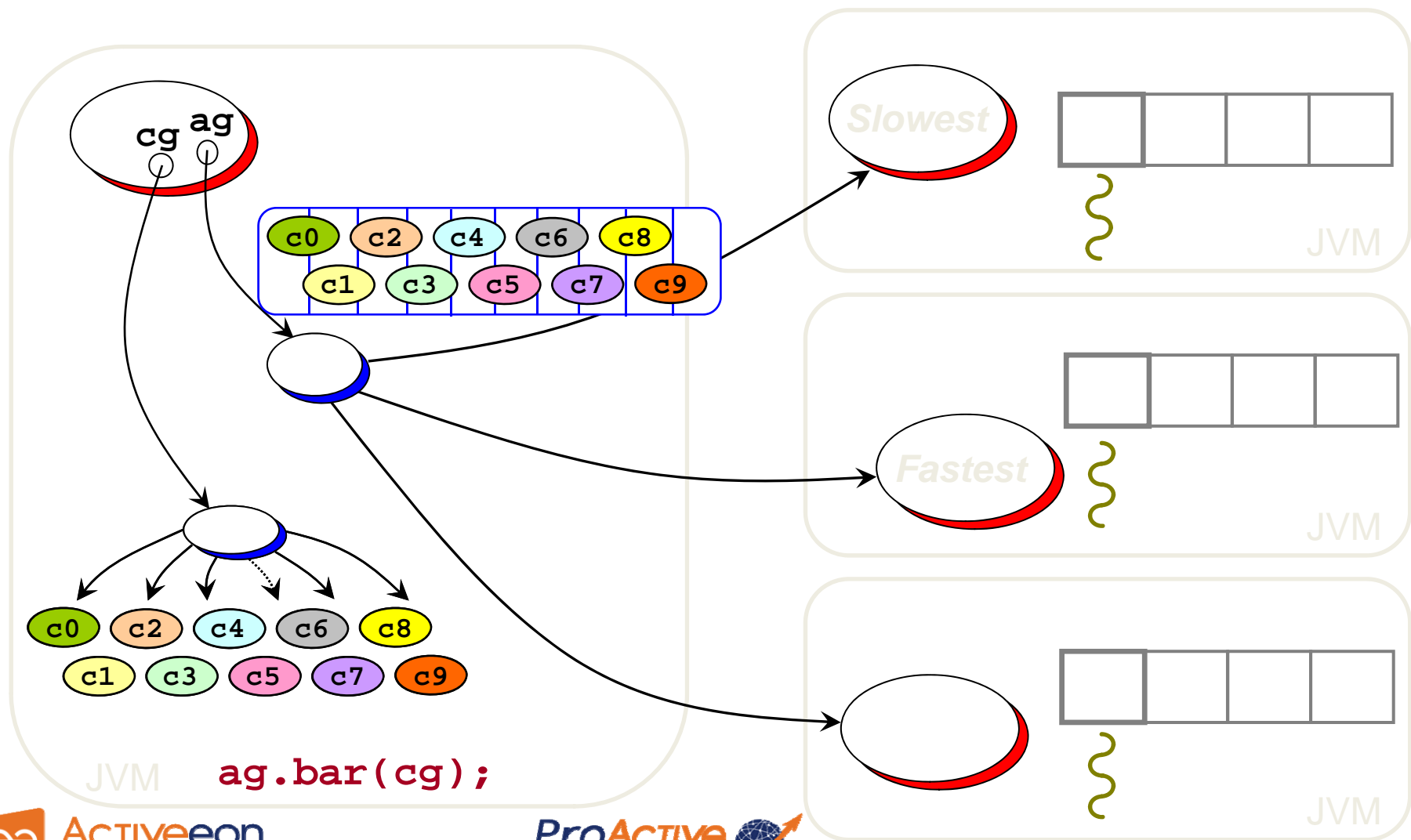
Broadcast and Scatter

Broadcast is the default behavior

Use a group as parameter, Scattered depends on rankings



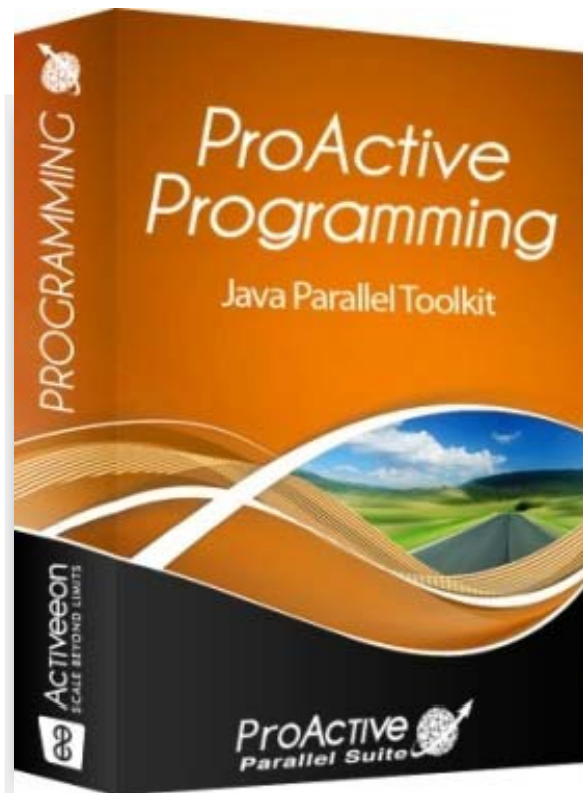
Dynamic Dispatch Group



Abstractions for Parallelism

The right Tool to do the Task right

ProActive Parallel Suite



- ☐ Workflows in Java
- ☐ Master/Workers
- ☐ SPMD
- ☐ Components
- ☐ ...

Core API
Active Objects
Asynchrony
Futures
Groups
Mobile Agents
MOP / AOP

Object-Oriented SPMD

Key Point

❑ **“MPI and programming languages from the 60’s will not make it”**

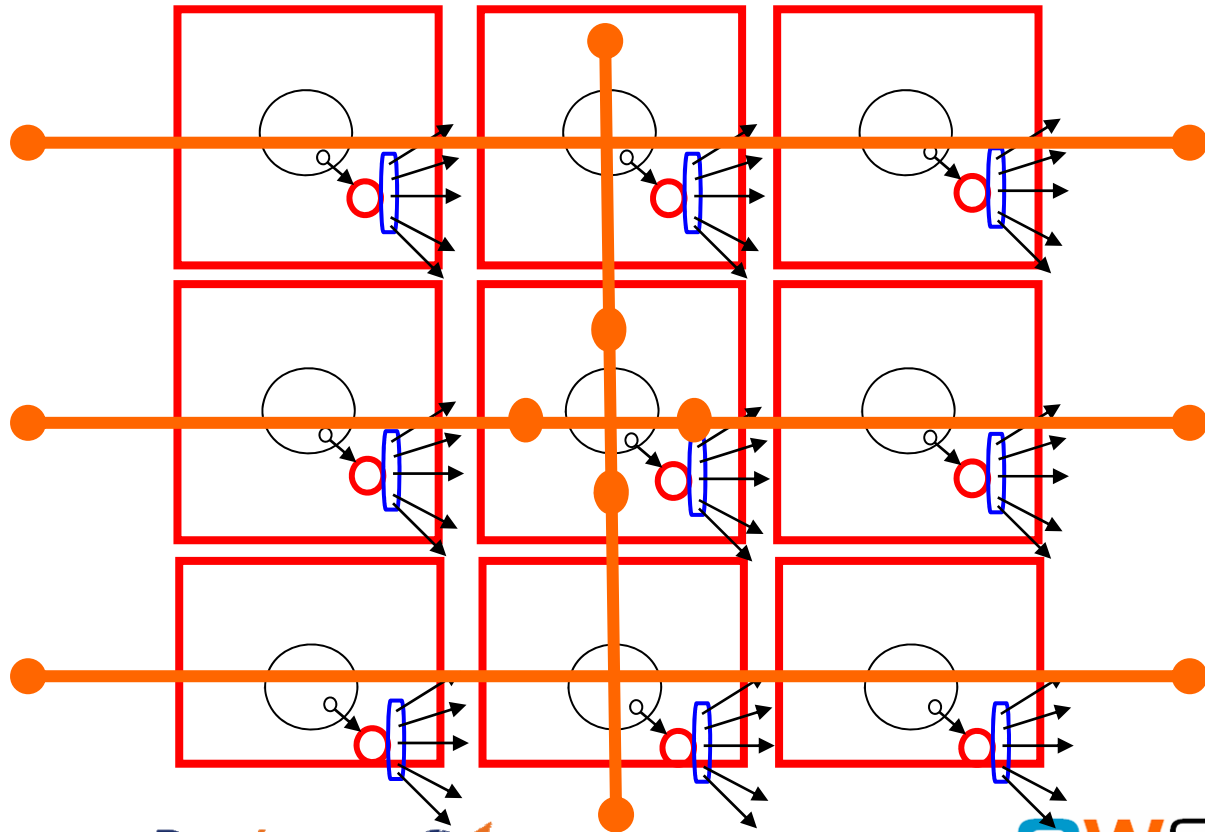
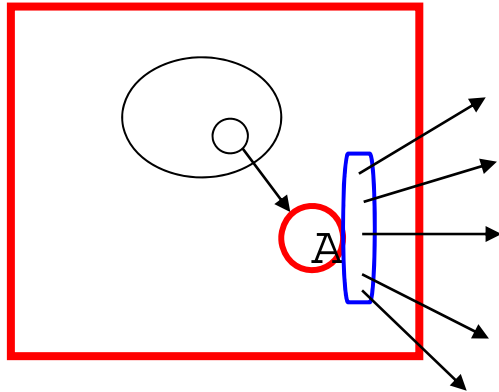
- **Jack Dongarra, 2/13/2009,**
- **Wake Forest University talk**

❑ **“It is time to get ride of MPI”**

- **Alan Edelman, MIT, 06/16/2010, Ter@tec**
- **ScilabTec’10 Users’ Day (30 mn. ago)**

OO SPMD: Object-Oriented SPMD

```
● A ag = newSPMDGroup ("A", [...], VirtualNode)
    // In each member
    ● myGroup.barrier ("2D"); // Global Barrier
    ● myGroup.barrier ("vertical"); // Any Barrier
    ● myGroup.barrier ("north","south","east","west");
```



IC2D: Optimizing

The screenshot displays the IC2D Monitoring View and Job Monitoring View. The Monitoring View shows a hierarchical tree of virtual nodes. The Job Monitoring View shows a detailed view of the DefaultVN job, including its sub-jobs and the nodes they are running on. The console at the bottom shows a message about a monitored node.

Monitoring View:

- Virtual nodes: ☒ Renderer, ☐ DefaultVN, ☐ Dispatcher, ☐ User
- PA_JVM1357457629_be...
 - Node Node60562498...
 - DinnerLayout#2
 - Table#3
 - Philosopher#4
 - Philosopher#5
 - Philosopher#6
 - Philosopher#7
 - Philosopher#8

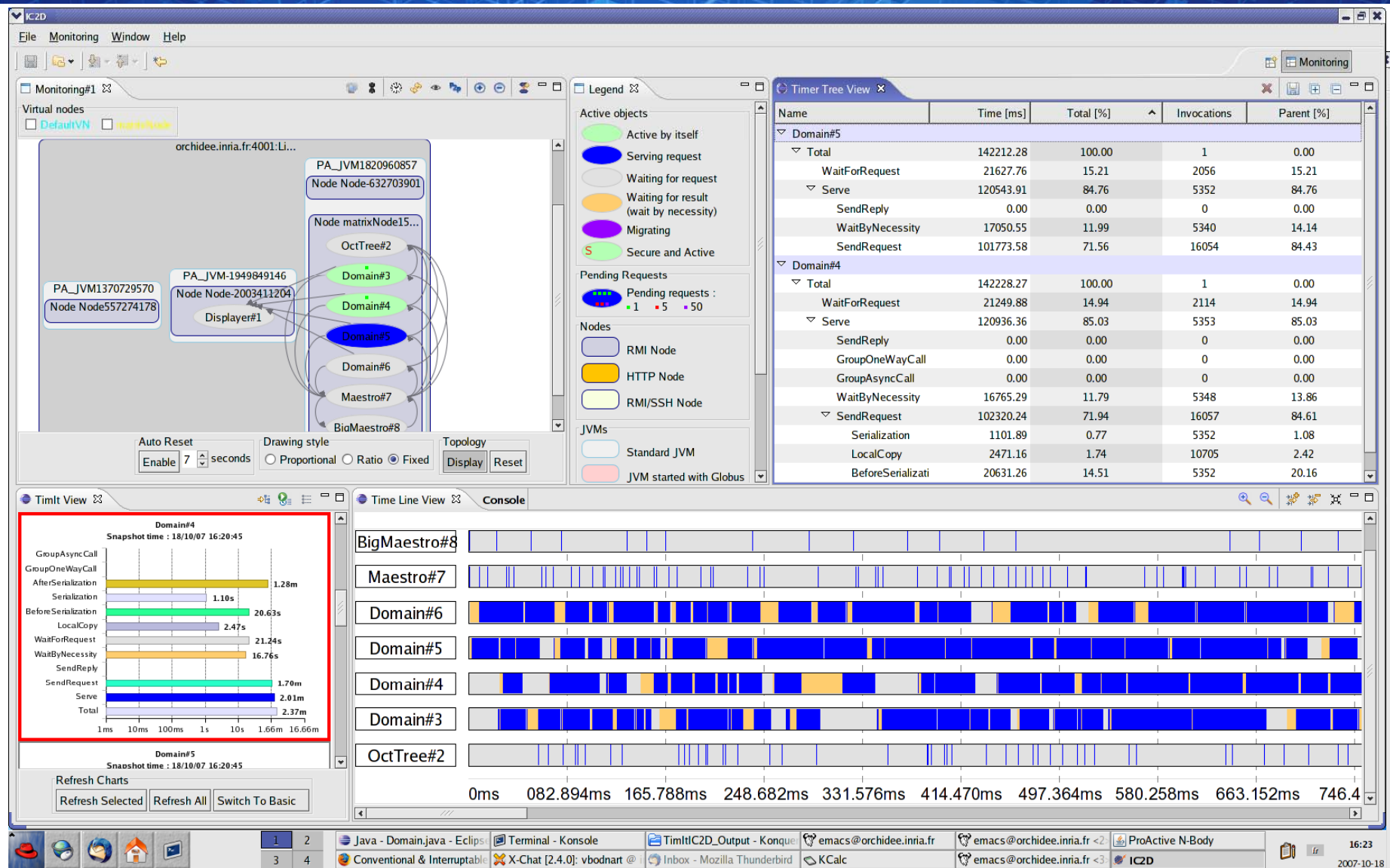
Job Monitoring View:

- DefaultVN (JOB-1357457629)
 - bebita.inria.fr:1099:OS un
 - PA_JVM1357457629_
 - Node Node60562498...
 - DinnerLayout#2
 - Table#3 (JOB-1357457629)
 - Philosopher#4 (JOB-1357457629)
 - Philosopher#5 (JOB-1357457629)
 - Philosopher#6 (JOB-1357457629)

Console:

```
15:09:15 => NodeObject id=Node-455186381 already monitored, check for new active objects
```

IC2D

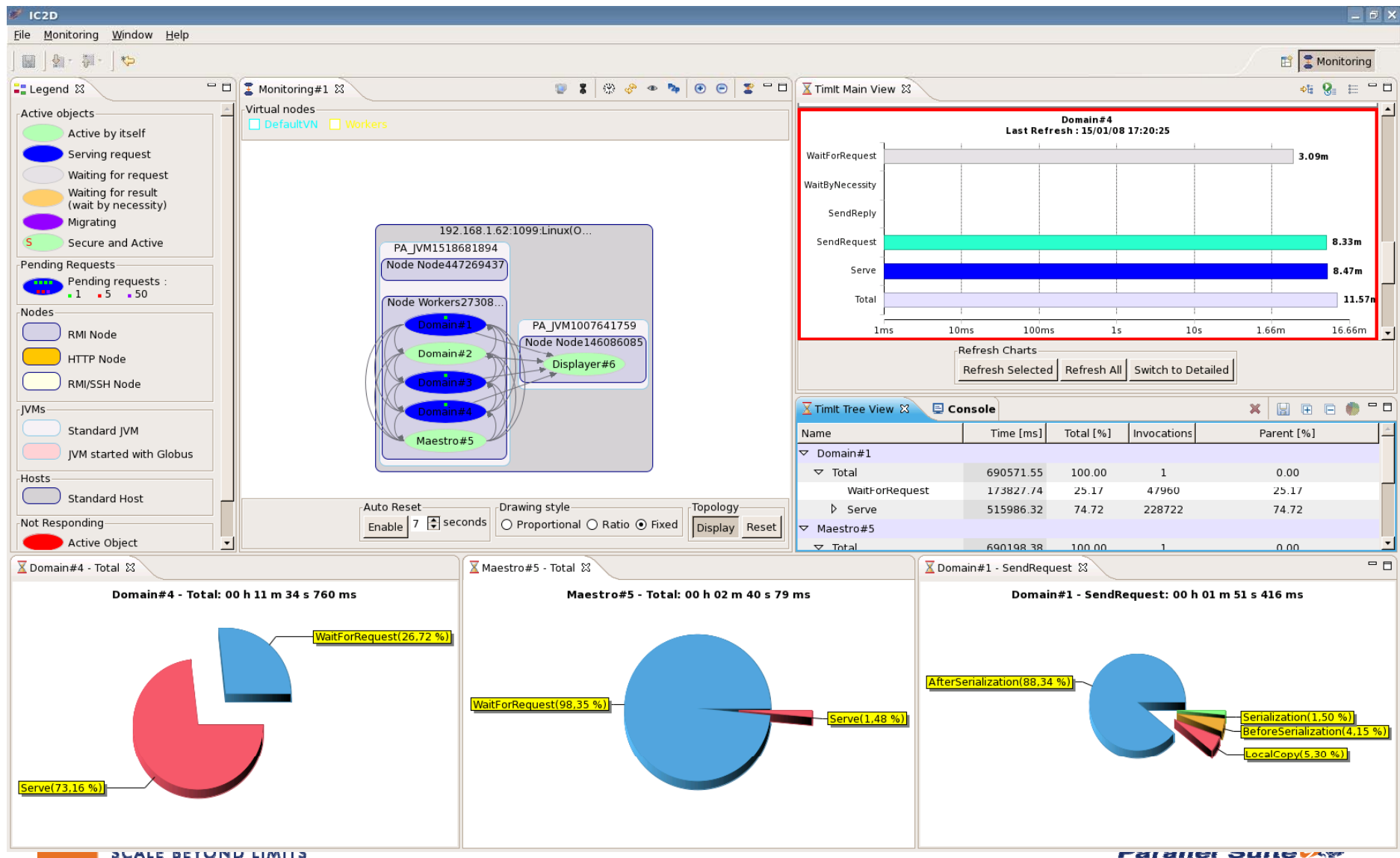


ChartIt



SCALE BEYOND LIMITS

Pies for Analysis and Optimization

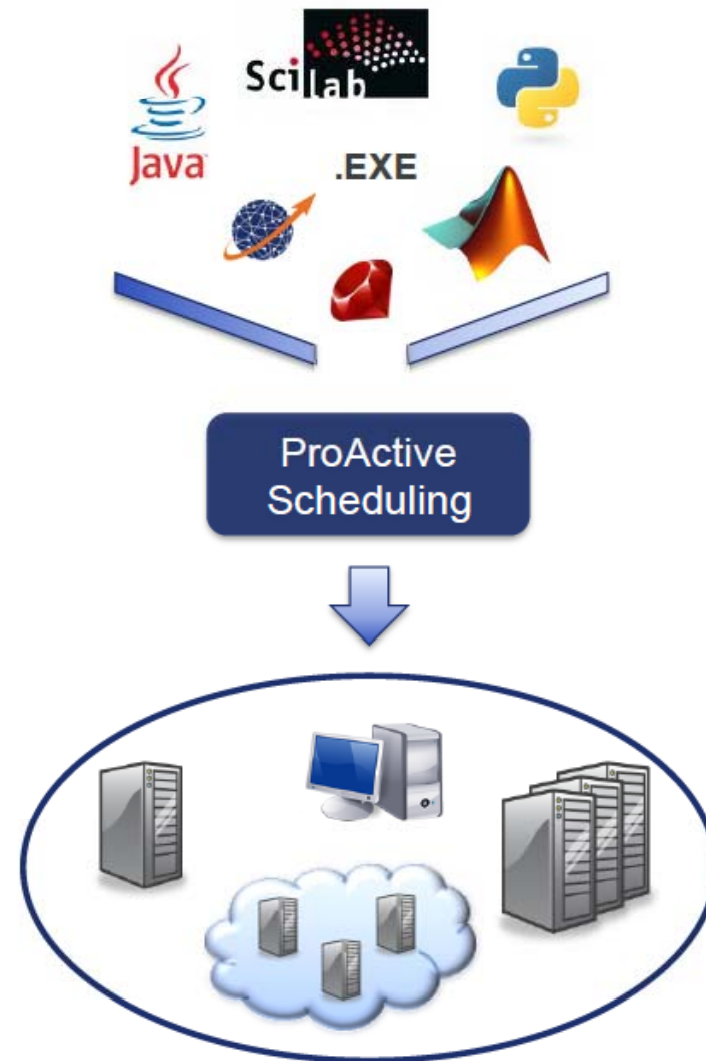
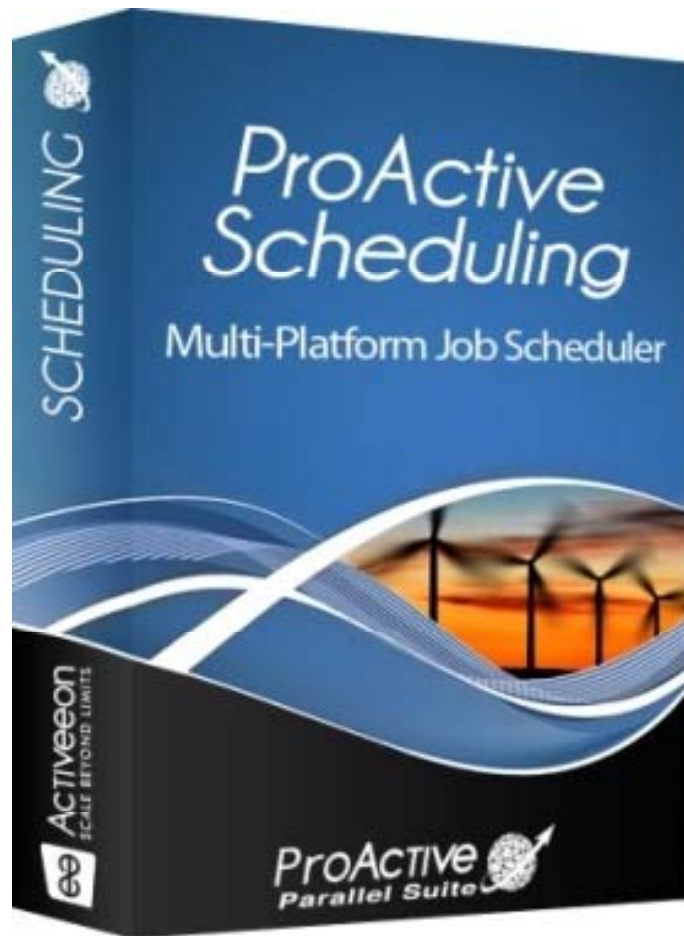


Video 1: IC2D Optimizing Monitoring, Debugging, Optimizing



Scheduling & Resourcing

ProActive Scheduling



ProActive Scheduling Big Picture

The screenshot displays the ProActive Scheduler interface. The top section shows three panels: Pending (674), Running (60), and Finished (31). The bottom section shows a detailed view of Job 2008, which has 8 tasks. The Job Info panel on the right provides details about the job, including its state (Pending), name (job_with_dep), priority (Normal), and task counts.

Pending (674)

Id	State	User	Priority	Name
1996	Pending	jl	Normal	job_with_dep
1997	Pending	jl	Normal	job_with_dep
1998	Pending	jl	Normal	job_with_dep
1999	Pending	jl	Normal	job_with_dep
2000	Pending	jl	Normal	job_with_dep
2001	Pending	jl	Normal	job_with_dep
2002	Pending	jl	Normal	job_with_dep
2003	Pending	jl	Normal	job_with_dep
2004	Pending	jl	Normal	job_with_dep
2005	Pending	jl	Normal	job_with_dep
2006	Pending	jl	Normal	job_with_dep
2007	Pending	jl	Normal	job_with_dep
2008	Pending	jl	Normal	job_with_dep
2009	Pending	jl	Normal	job_with_dep
2010	Pending	jl	Normal	job_with_dep

Running (60)

Id	State	Progress	# Finished	User	Priority
1313	Running	4/8	4/8	user1	Normal
1314	Running	4/8	4/8	user1	Normal
1315	Running	7/8	7/8	admin	Normal
1316	Running	4/8	4/8	user1	Normal
1317	Running	7/8	7/8	admin	Normal
1318	Running	4/8	4/8	user1	Normal
1319	Running	7/8	7/8	admin	Normal
1320	Running	3/8	3/8	user1	Normal
1321	Running	7/8	7/8	admin	Normal
1322	Running	3/8	3/8	user1	Normal
1323	Running	7/8	7/8	admin	Normal
1324	Running	2/8	2/8	user1	Normal
1325	Running	2/8	2/8	user1	Normal
1326	Running	2/8	2/8	user1	Normal
1327	Running	2/8	2/8	user1	Normal

Finished (31)

Id	State	User	Priority	Name
010	Finished	jl	Low	job_proActive
008	Finished	jl	Low	job_proActive
005	Finished	jl	Low	job_proActive
001	Finished	jl	Low	job_proActive
006	Finished	jl	Low	job_proActive
004	Finished	jl	Low	job_proActive
003	Finished	jl	Low	job_proActive
009	Finished	jl	Low	job_proActive
007	Finished	jl	Low	job_proActive
002	Finished	jl	Low	job_proActive
245	Finished	user1	Normal	job_with_dep
246	Finished	user1	Normal	job_with_dep
247	Finished	user1	Normal	job_with_dep
252	Finished	admin	Normal	job_with_dep
253	Finished	admin	Normal	job_with_dep

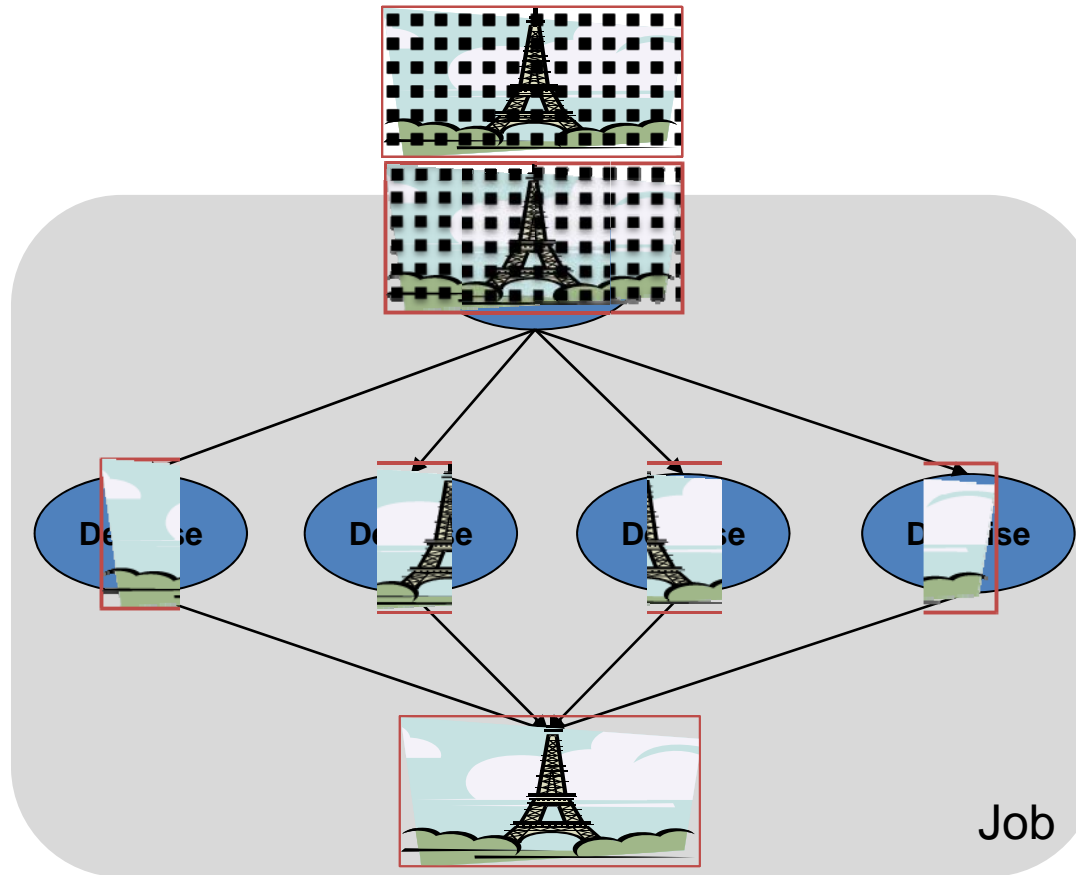
Job 2008 has 8 tasks

Id	State	Name	Host name	Start time	Finished time	Re-run	Description
200800: Submitted	task4	n/a	Not yet	Not yet	0/2	This task will sleep 5s	
200800: Submitted	task2	n/a	Not yet	Not yet	0/1	This task will sleep 10s	
200800: Submitted	task6	n/a	Not yet	Not yet	0/1	This task will sleep 8s	
200800: Submitted	task1	n/a	Not yet	Not yet	0/2	This task will sleep 6s	
200800: Submitted	task5	n/a	Not yet	Not yet	0/1	This task will sleep 2s	
200800: Submitted	task7	n/a	Not yet	Not yet	0/2	This task will sleep 6s	
200800: Submitted	task3	n/a	Not yet	Not yet	0/1	This task will sleep 4s	
200800: Submitted	task8	n/a	Not yet	Not yet	0/1	This task will sleep 6s	

Job Info

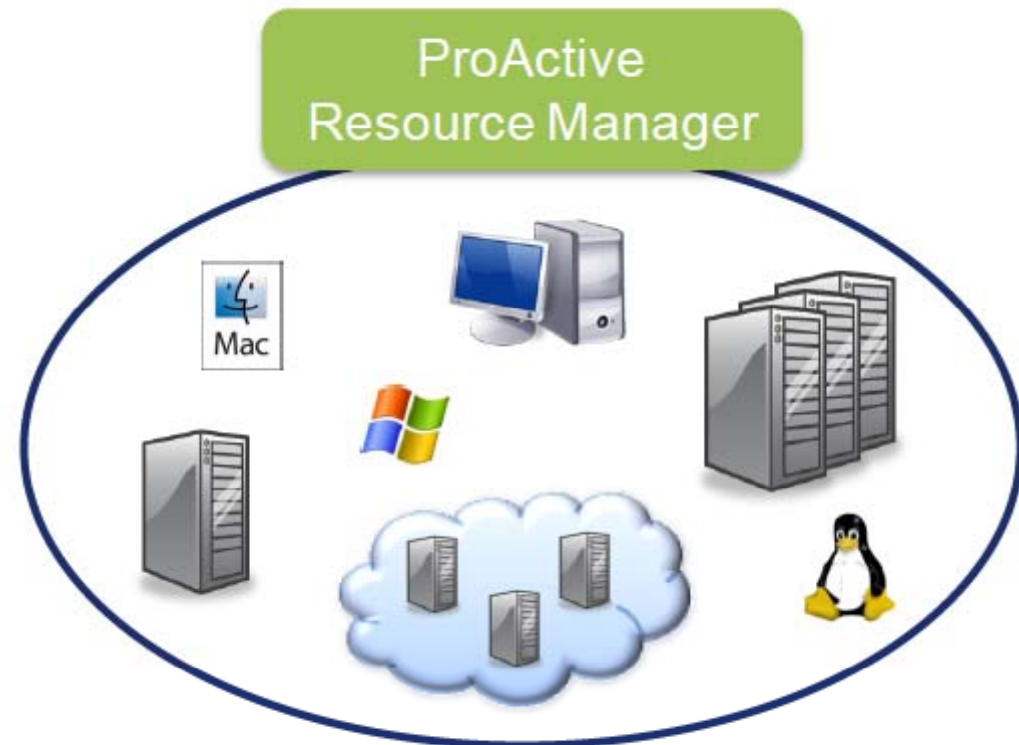
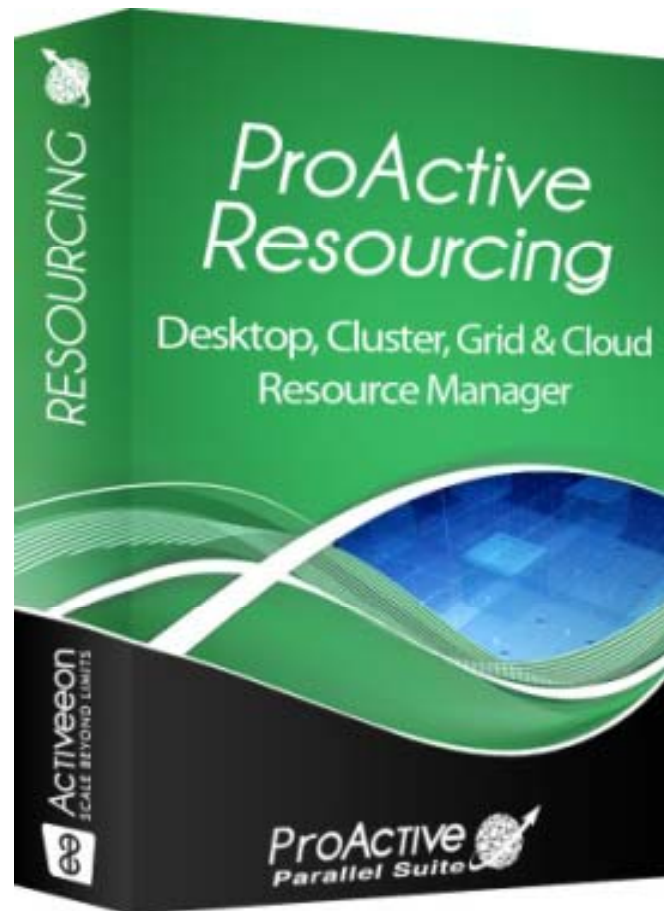
Property	Value
Id	2008
State	Pending
Name	job_with_dep
Priority	Normal
Pending tasks number	0
Running tasks number	0
Finished tasks number	0
Total tasks number	8
Submitted time	09:40:06 03/12/08
Started time	Not yet
Finished time	Not yet

Workflow Example : Picture Denoising

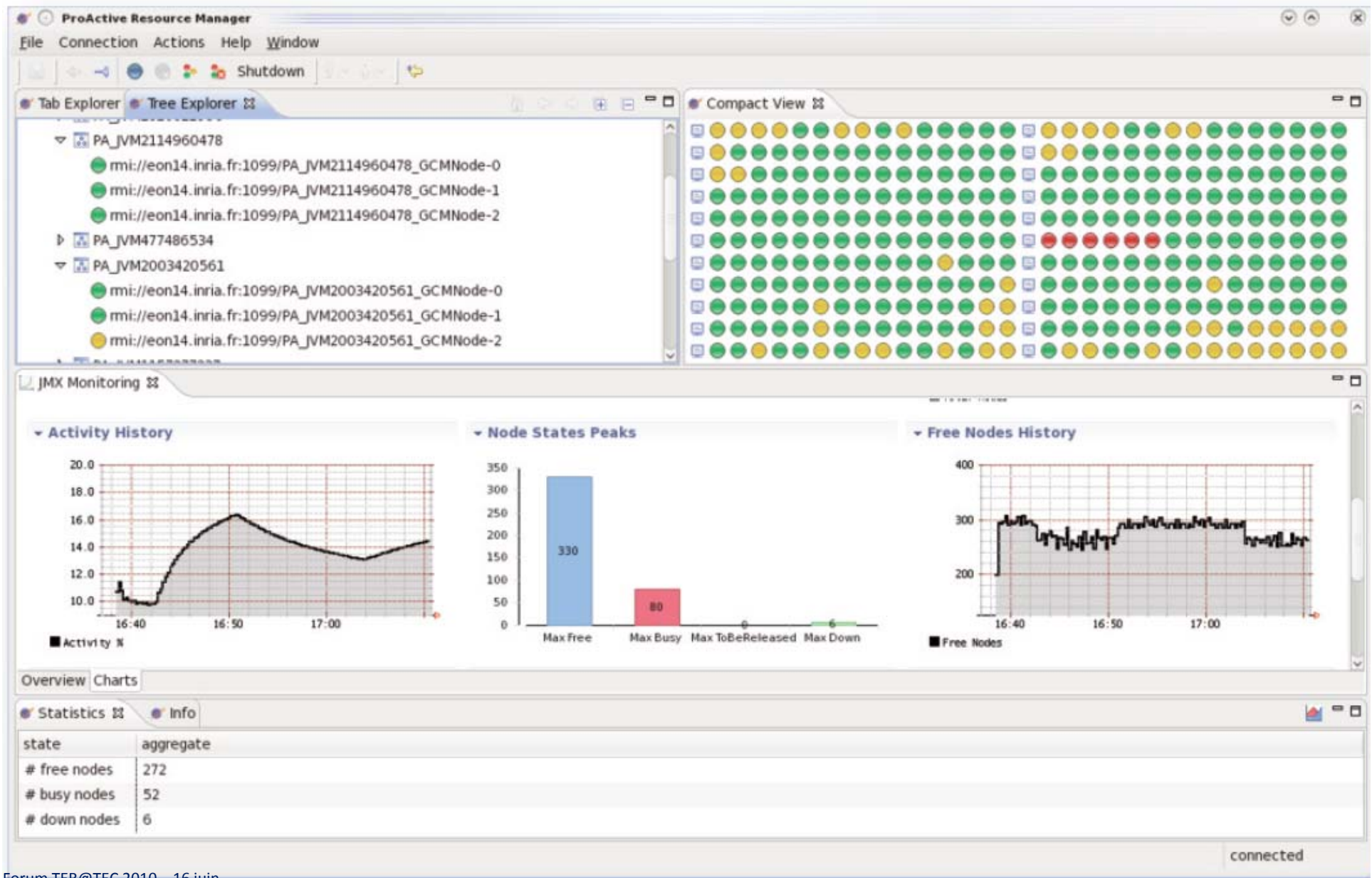


- with **selection** on native executable availability (ImageMagik, GREYstoration)
 - Multi-platform selection and command generation
- with **file transfer** in pre/post scripts

ProActive Resourcing

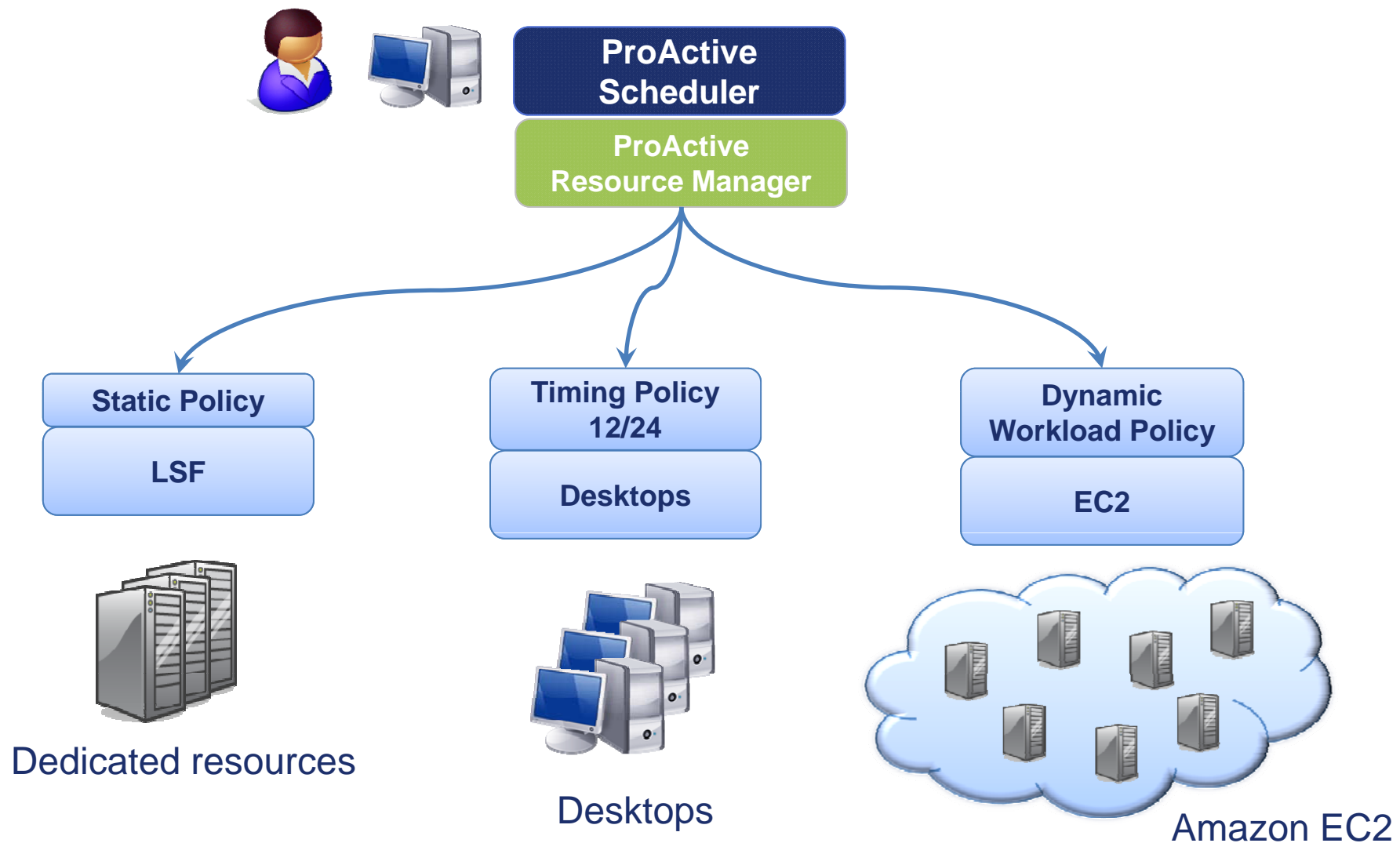


RESOURCING User Interface



Clusters to Grids to Clouds e.g. on Amazon EC2

Node source Usecase : Configuration for external cloud with EC2



Video 2: Scheduler, Resource Manager

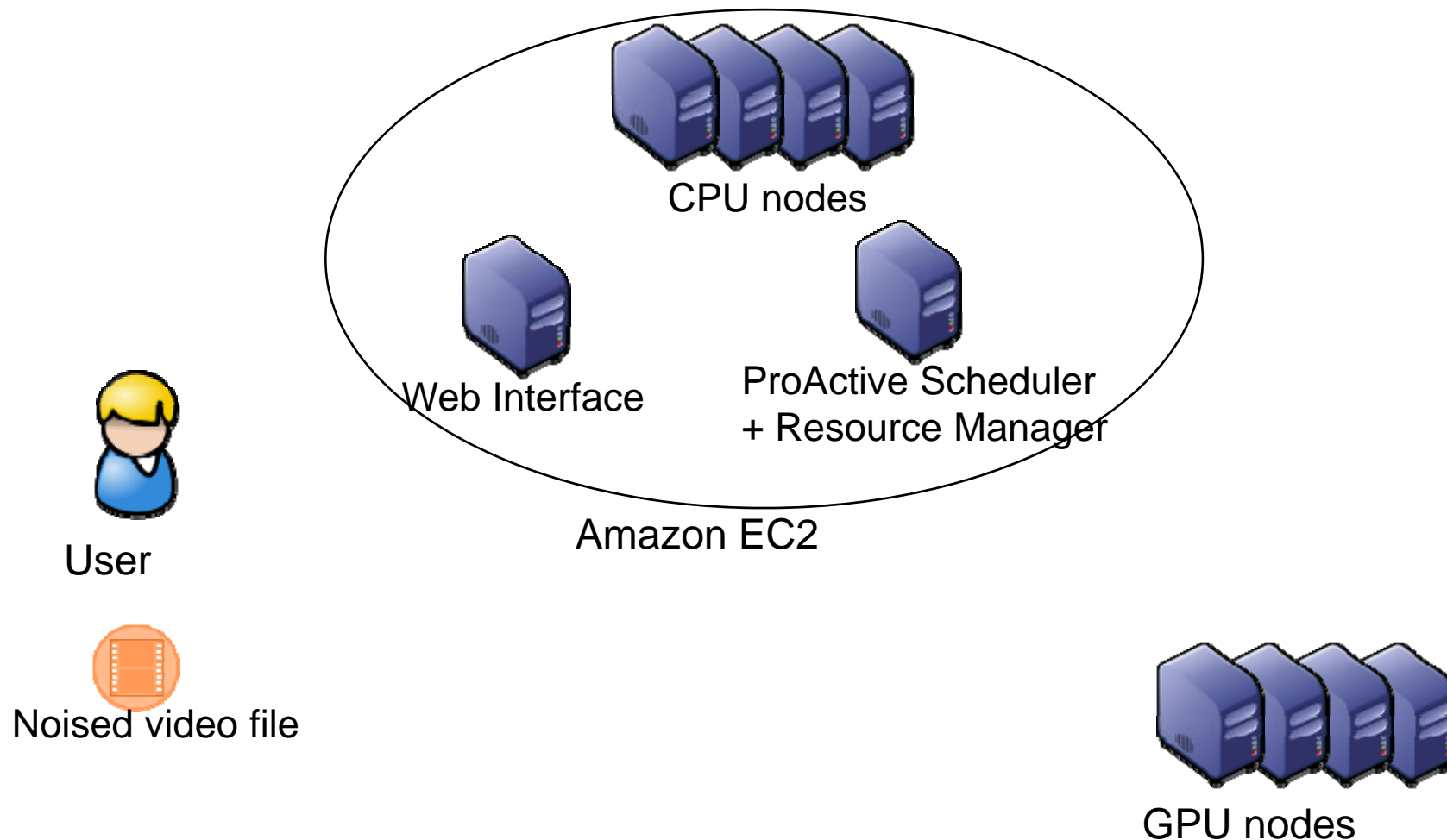


3. Cloud Seeding with GPU

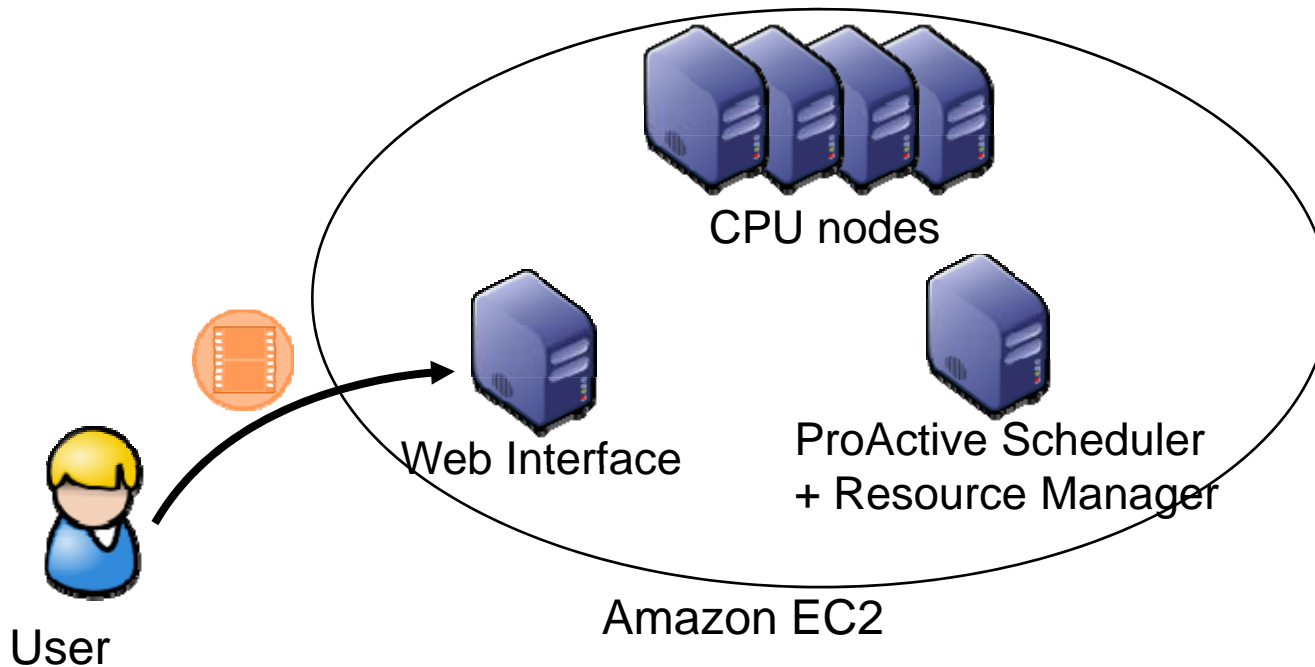
Cloud Seeding with ProActive

- ❑ Amazon EC2 Execution
- ❑ *Cloud Seeding* strategy to mix heterogeneous computing resources :
 - External GPU resources

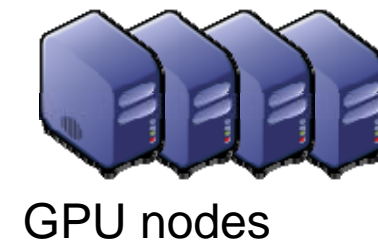
Cloud Seeding with ProActive



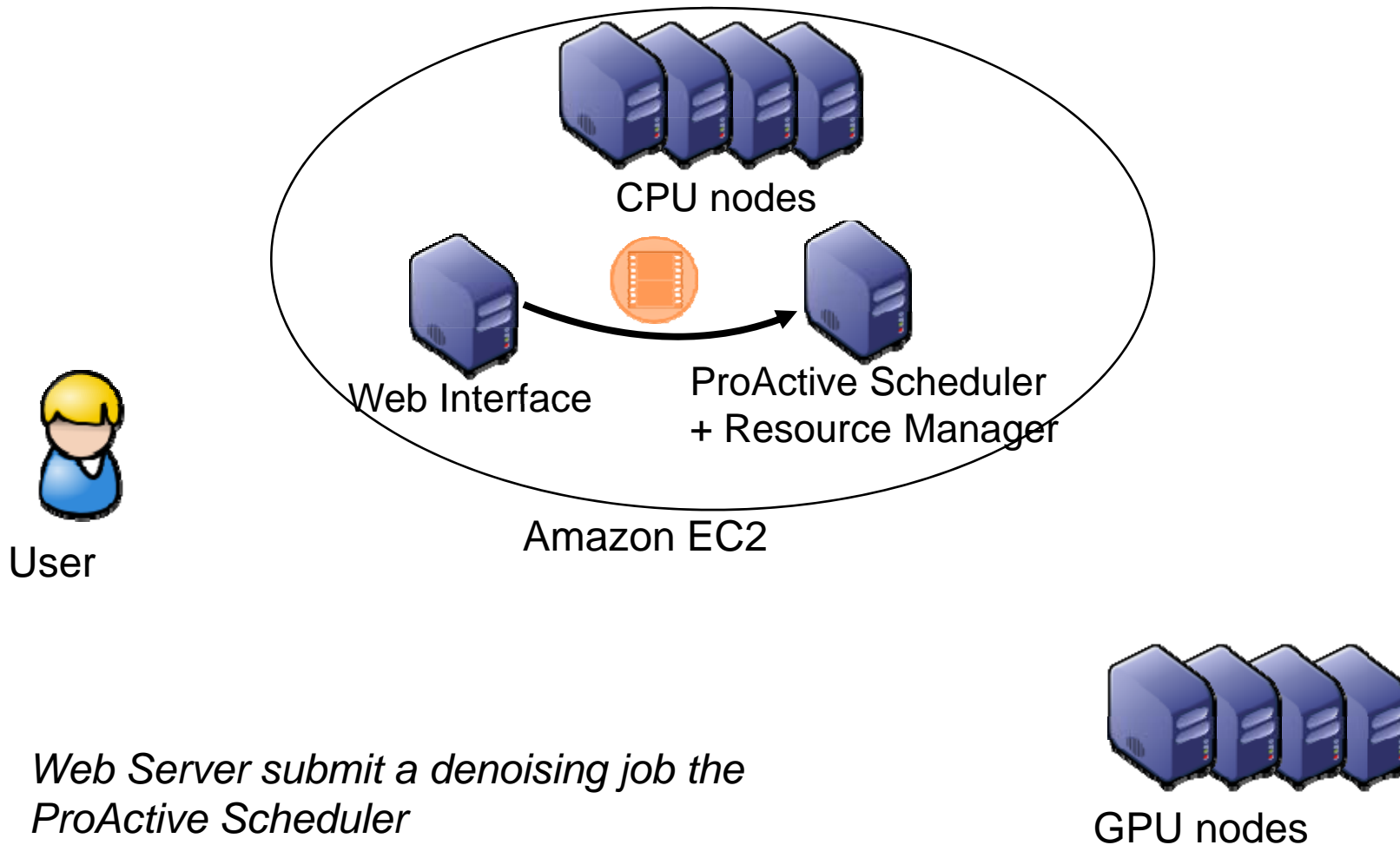
Cloud Seeding with ProActive



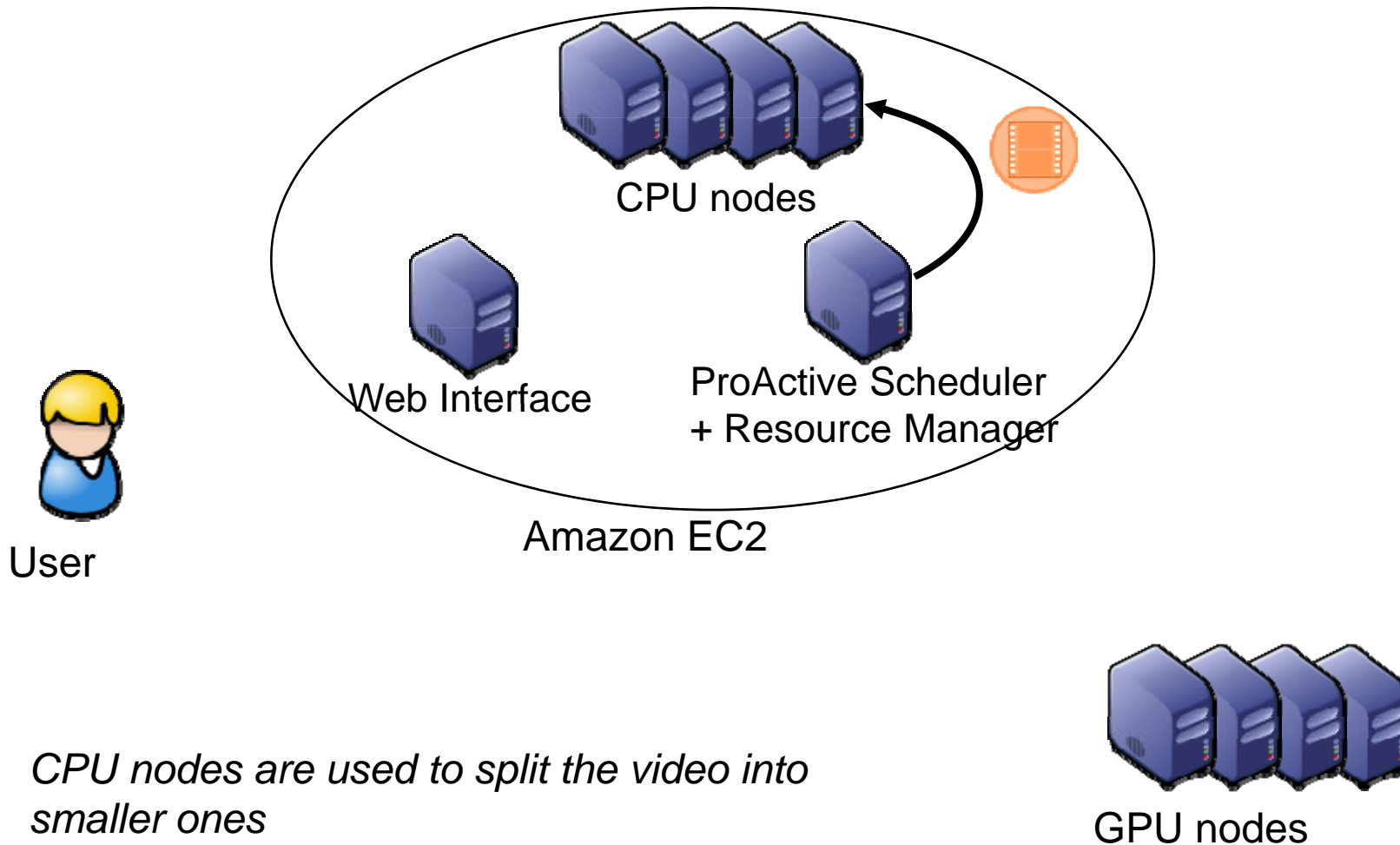
User submit its noised video to the web interface



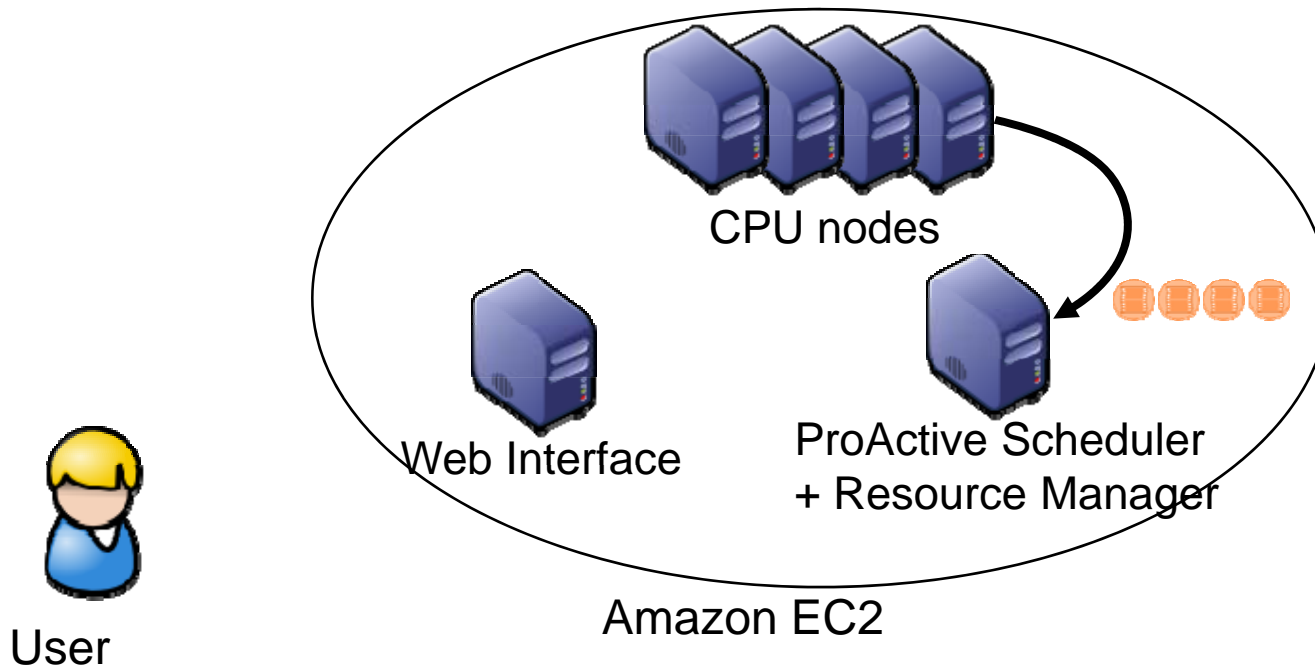
Cloud Seeding with ProActive



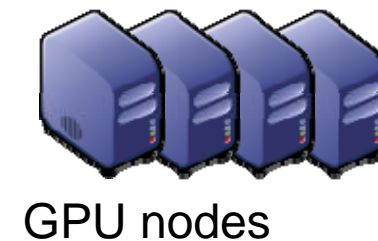
Cloud Seeding with ProActive



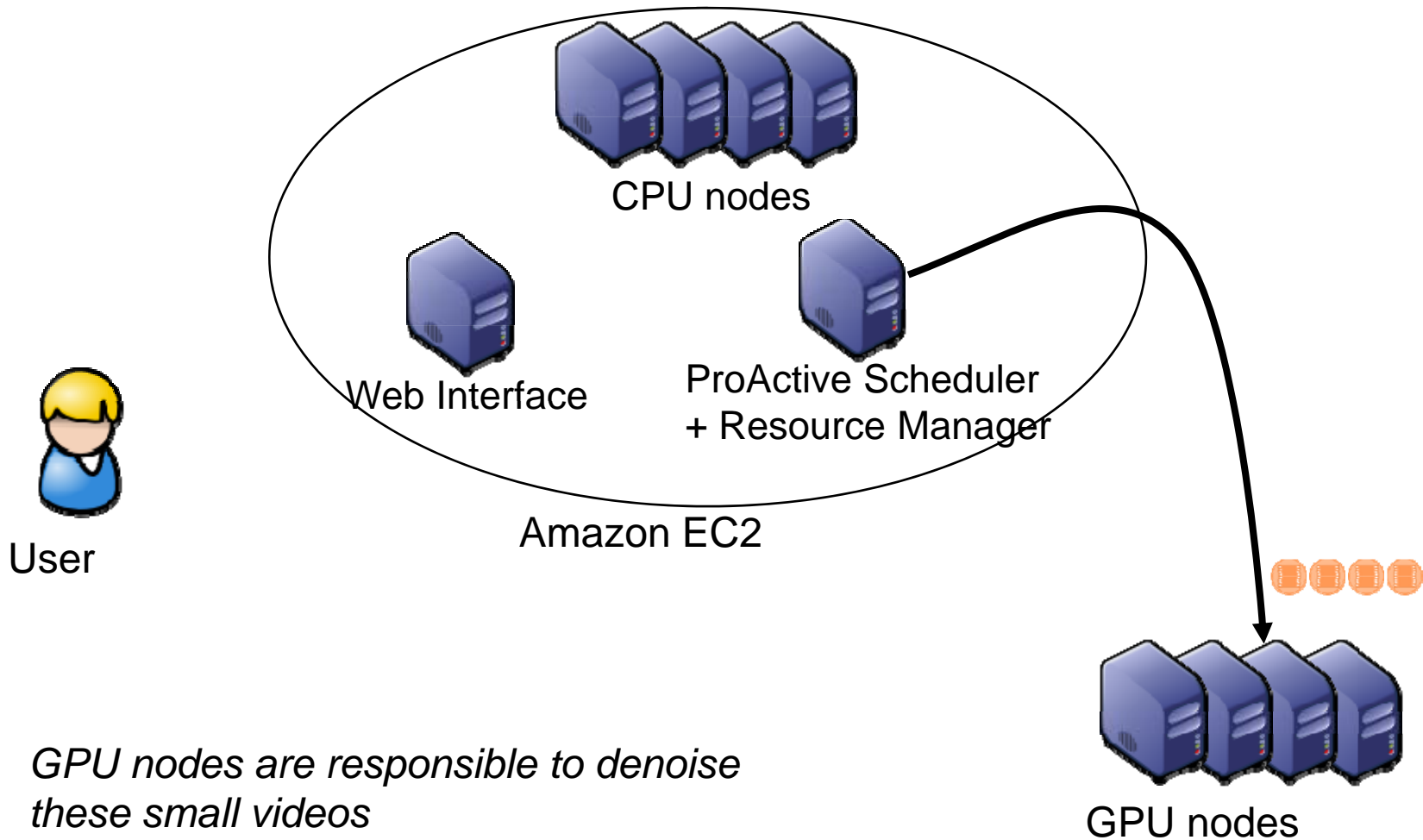
Cloud Seeding with ProActive



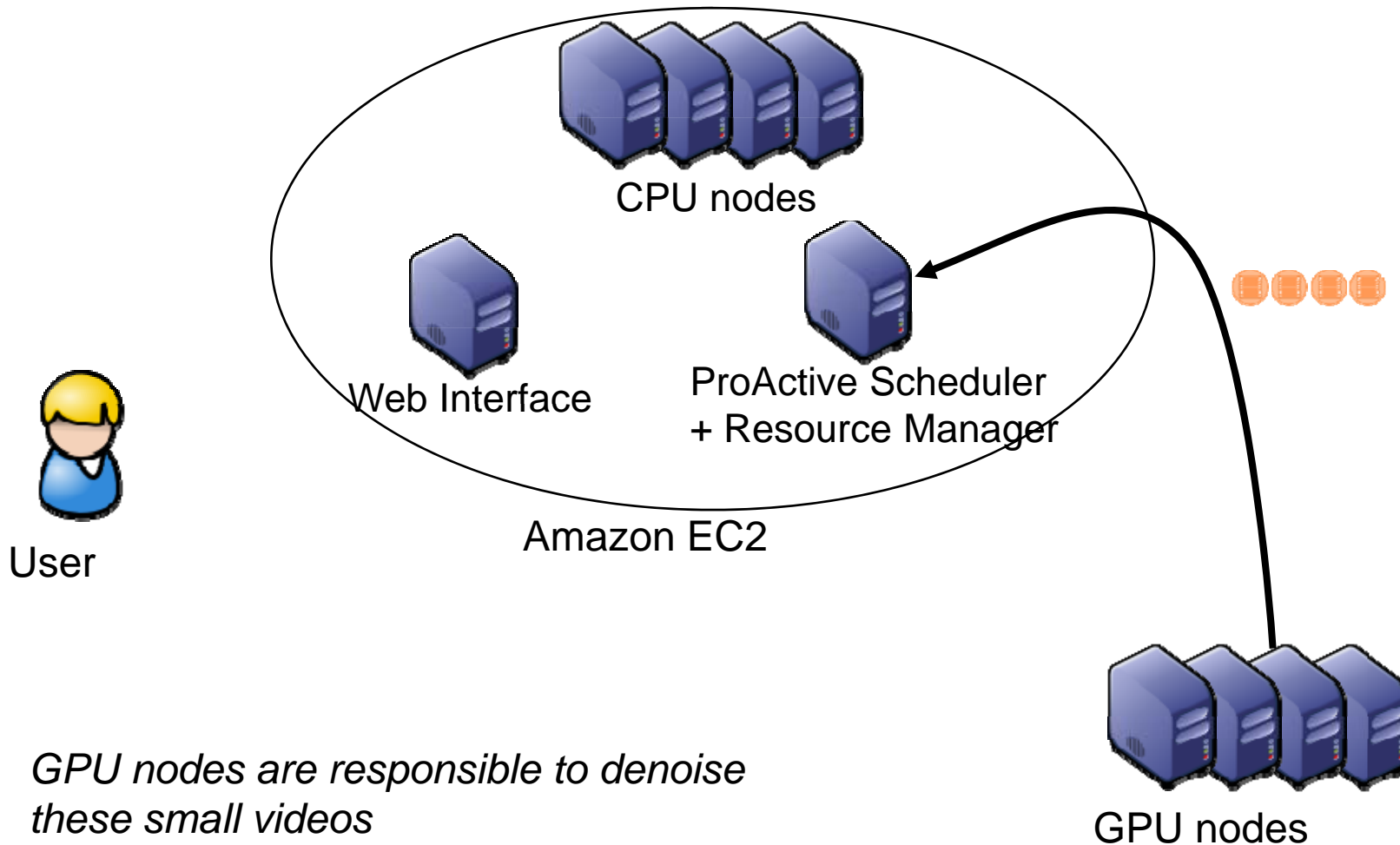
CPU nodes are used to split the video into smaller ones



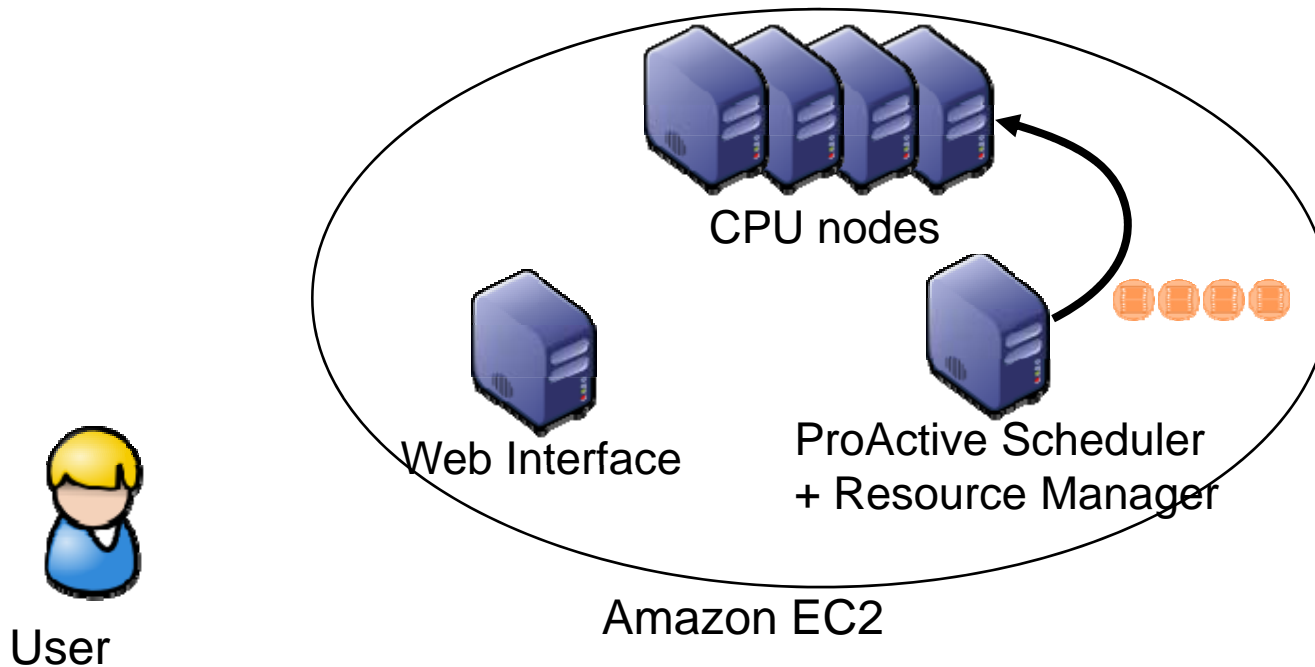
Cloud Seeding with ProActive



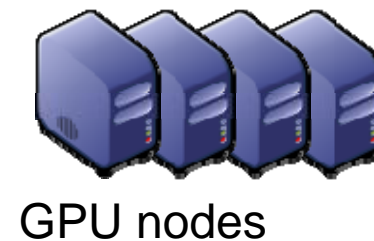
Cloud Seeding with ProActive



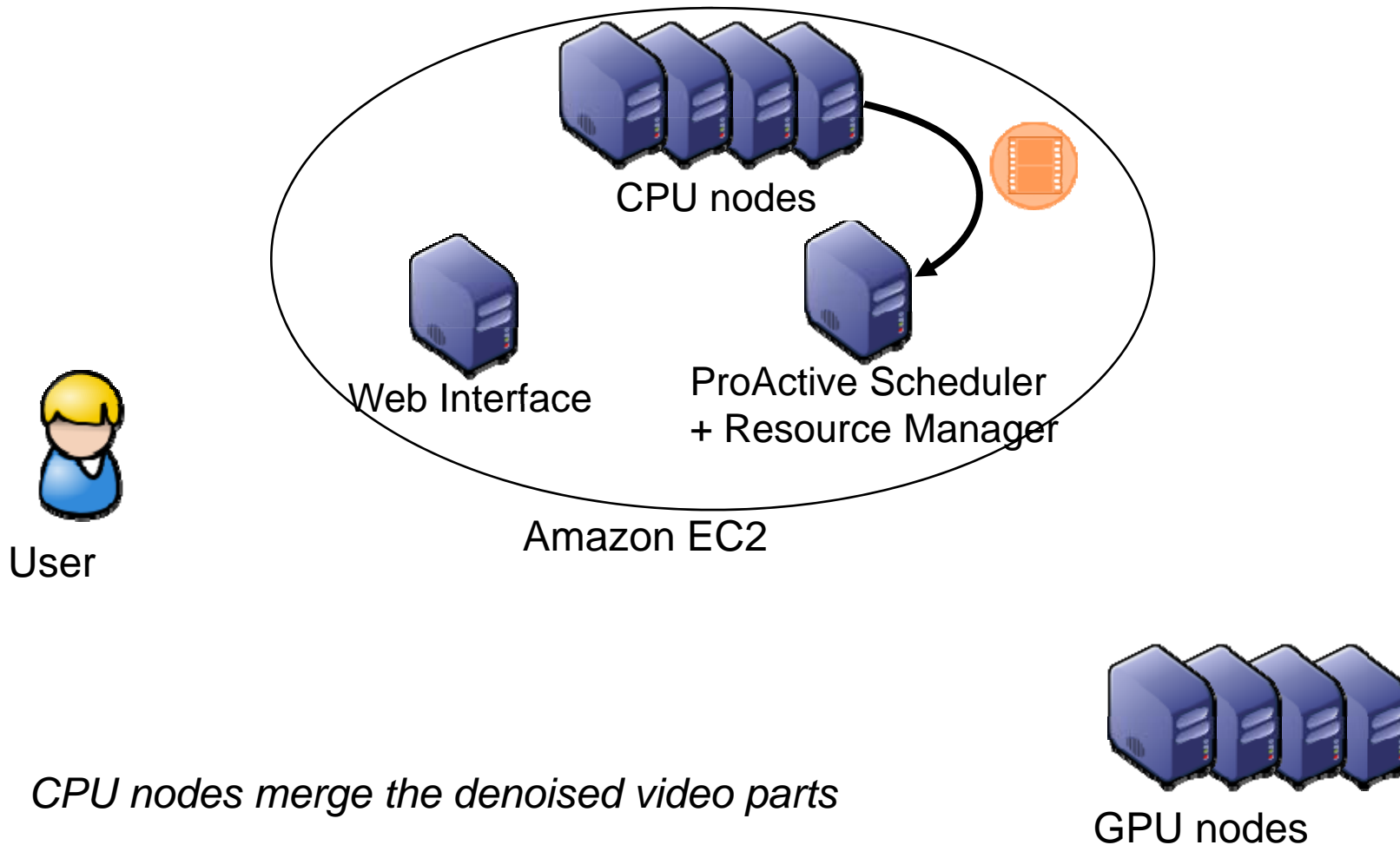
Cloud Seeding with ProActive



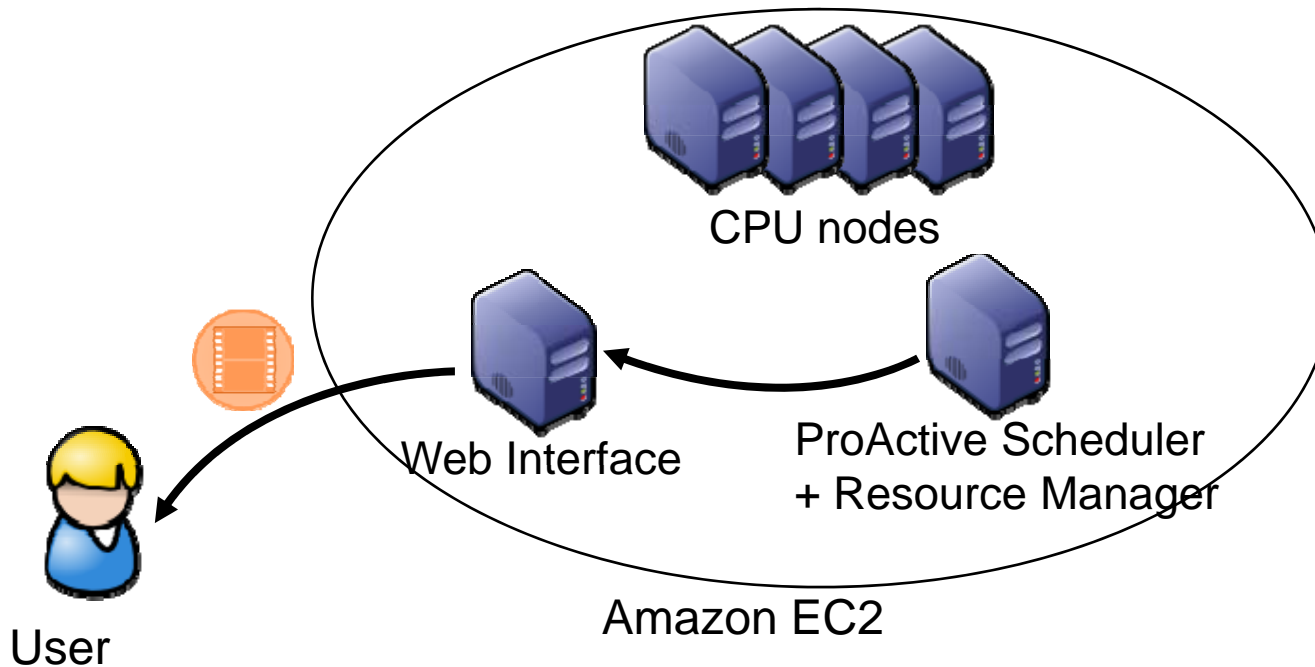
CPU nodes merge the denoised video parts



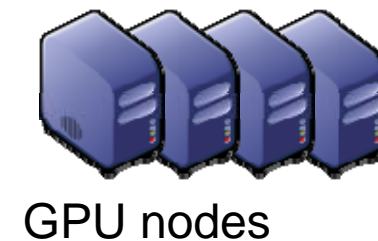
Cloud Seeding with ProActive



Cloud Seeding with ProActive



The final denoised video is sent back to the user



4. ProActive PACA GRID: Cloud Portal with GPUs

Main

- » [Welcome](#)
- » [Monitor](#)
- » [Nodes configuration](#)
- » [Download](#)
- » [Tutorials](#)
- » [links](#)



Welcome

Welcome to ProActive PACA Grid web site

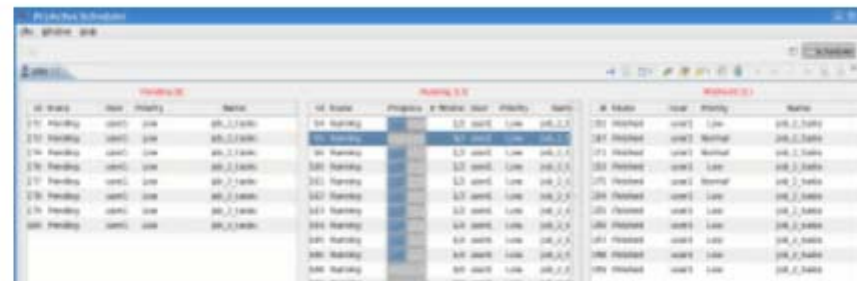
ProActive PACA Grid is a set of machines accessible via Graphical Interactive interfaces based on ProActive Parallel Suite (<http://proactive.inria.fr>). The machines are currently deployed within INRIA Sophia Antipolis networks. The Cloud aggregates dedicated machines, both Linux and Windows, and spare desktop machines, dynamically added during nights and week-ends. This Grid is available for INRIA and UNSA members that need to accelerate their scientific applications. Upon request, other PACA labs and SMEs can also access the ProActive PACA Grid.

In production today : Download Graphical client in [download](#) page and schedule your jobs !

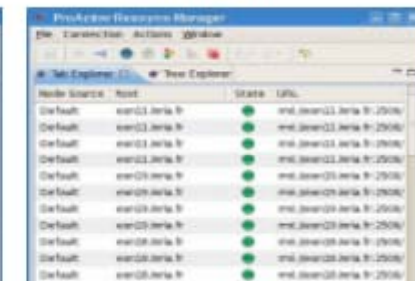
On going : Direct Web Access to CPER Cloud cluster by Java Web Start :

[Web Access to the ProActive Scheduler :](#)

[Web Access to the ProActive Resource Manager :](#)



Node	Source	Host	Port	State	Priority	Usage
100	ProActive	192.168.1.1	8080	Ready	High	0%
101	ProActive	192.168.1.2	8080	Ready	High	0%
102	ProActive	192.168.1.3	8080	Ready	High	0%
103	ProActive	192.168.1.4	8080	Ready	High	0%
104	ProActive	192.168.1.5	8080	Ready	High	0%
105	ProActive	192.168.1.6	8080	Ready	High	0%
106	ProActive	192.168.1.7	8080	Ready	High	0%
107	ProActive	192.168.1.8	8080	Ready	High	0%
108	ProActive	192.168.1.9	8080	Ready	High	0%
109	ProActive	192.168.1.10	8080	Ready	High	0%



Node	Source	Host	Port	State	Usage
100	ProActive	192.168.1.1	8080	Ready	0%
101	ProActive	192.168.1.2	8080	Ready	0%
102	ProActive	192.168.1.3	8080	Ready	0%
103	ProActive	192.168.1.4	8080	Ready	0%
104	ProActive	192.168.1.5	8080	Ready	0%
105	ProActive	192.168.1.6	8080	Ready	0%
106	ProActive	192.168.1.7	8080	Ready	0%
107	ProActive	192.168.1.8	8080	Ready	0%
108	ProActive	192.168.1.9	8080	Ready	0%
109	ProActive	192.168.1.10	8080	Ready	0%

The ProActive PACA Grid Platform (4)

Total:

- ❑ 816 Cores
- ❑ 480 CUDA Cores
- ❑ 14.8TB Storage

Publically Available Today



ProActive in Cloud Stack



ProActive
Cloud Portal

Mosso
Google App Engine
Rails One

Salesforce
Gmail
Gliffy

Joyent
Amazon Web Svcs
Nirvanix
XCalibre
Akamai

PaaS

SaaS

IaaS

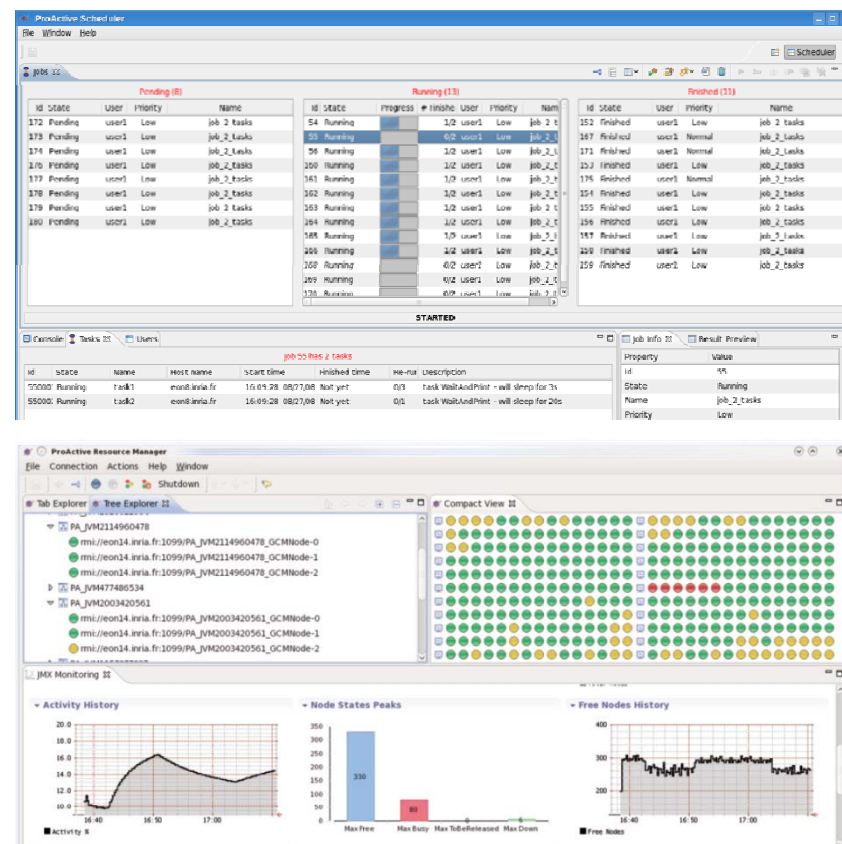
Cloud Computing

Utility Computing

Grid Computing

Cluster Computing

Super Computing

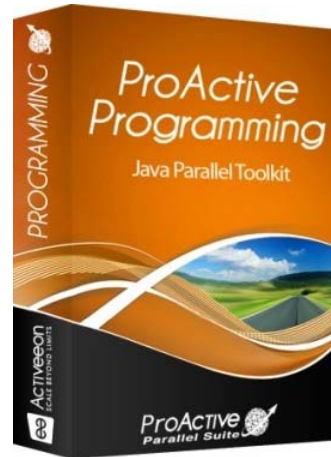


Conclusion

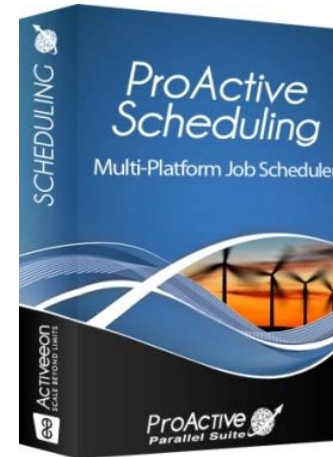
Conclusion



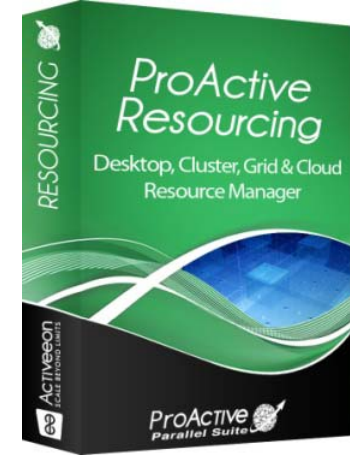
- Portability:
Windows, Linux, Mac
- Versatility:
Desktops, Grids, Clouds



Java Parallel
Toolkit



Multi-Platform
Job Scheduler



Resource
Manager

Free Professional
Open Source Software

OW²
Consortium

Manycores, Distribution and Cloud

Multi-Core: Ready for a strong evolution

Cloud: Smooth transition available (Desktop , Server, Cluster)

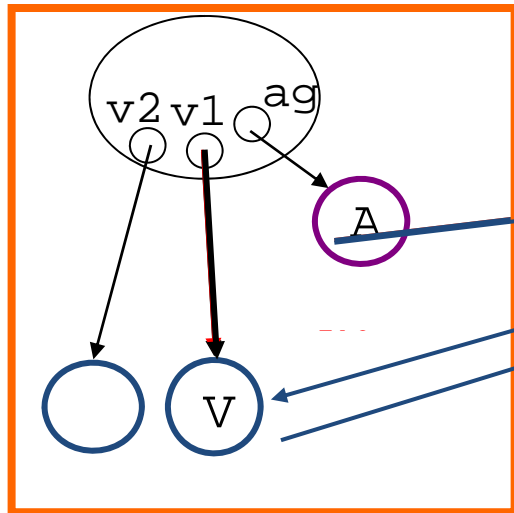


ProActive : Active objects

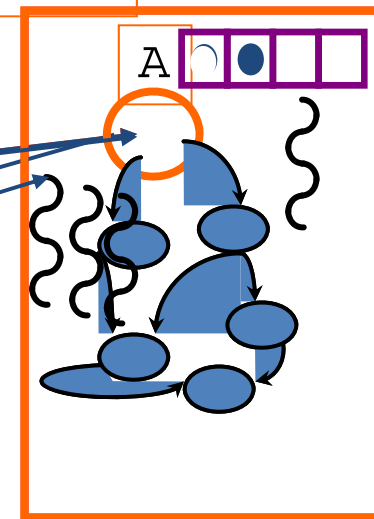
```

● A ag = newActive ("A", [...], VirtualNode)
● V v1 = ag.foo (param);
● V v2 = ag.bar (param);
...
● v1.bar(); //Wait-By-Necessity
    
```

JVM



JVM

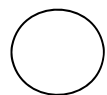


Wait-By-Necessity

is a

Dataflow

Synchronization



Java Object



Active Object



Req. Queue



Future Object



Proxy



Request



Thread



Activeeon
SCALE BEYOND LIMITS

ProActive
Parallel Suite

Use Case 1: Genomics

Resources set up

SOLID
machine from
AB Applied Biosystems



16
nodes



PBS

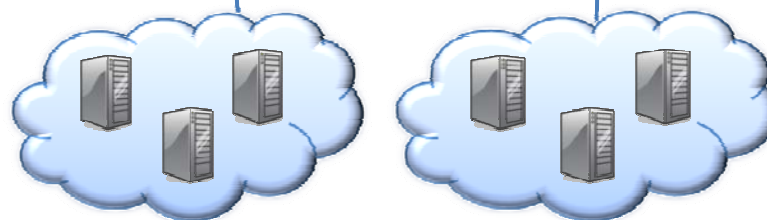
ProActive
Parallel Suite

Cluster



Desktops

**Nodes
can be
dynamically
added!**



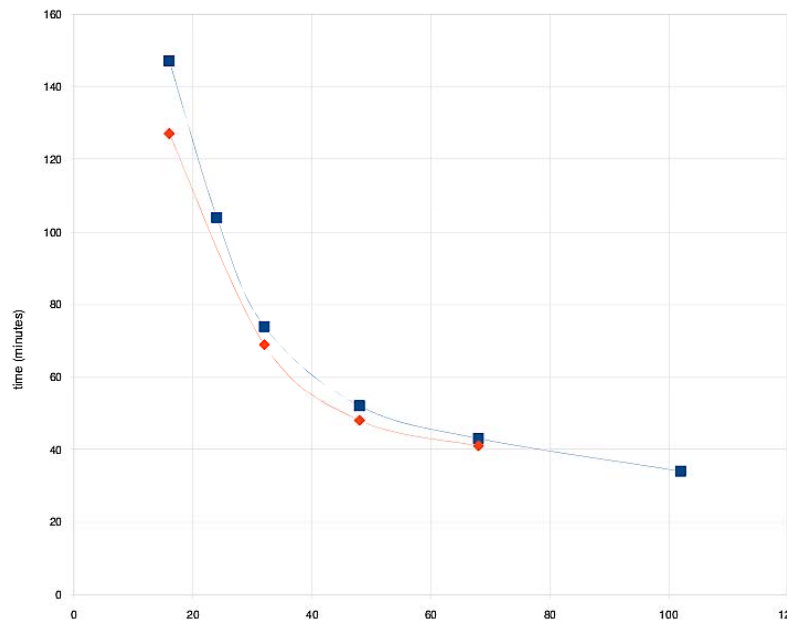
amazon
web services™

EC2

Clouds

First Benchmarks

- ❑ The distributed version with ProActive of Mapreads has been tested on the INRIA cluster with two settings: the Reads file is split in either 30 or 10 slices
- ❑ Use Case: Matching 31 millions Sequences with the Human Genome (M=2, L=25)



4 Time FASTER from 20 to 100
Speed Up of 80 / Th.
Sequential : 50 h → 35 mn

EC2 only test: nearly the same
performances as the local
SOLiD cluster (+10%)

**For only \$3,2/hour, EC2 has nearly the same perf. as
the local SOLiD cluster (16 cores, for 2H30)**

Benchmark: local vs. hybrid cloud

Use case: 3 runs performed in parallel containing a total of 28,5 millions of reads to be matched against the human genome

- **SOLID nodes only**

Standard configuration
using SOLiD embedded
nodes: 12



- Total computation time:
12.5 hours

- **SOLID and EC2 nodes**

12 SOLiD nodes
12 EC2 machines
(type: “m1large”, 2 nodes each)



- Total computation time:
8 hours



Gain: 4,5 hours (36% faster)
EC2 costs: \$40

Benchmark: local vs. EC2 cloud

	Execution time (min)	Cost (US\$)
Standard PBS config	300	NA
ProActive Amazon EC2	340	20 US\$

For only \$3,2/hour, the EC2 setup has nearly the same performances as the local SOLiD cluster

UC 2: Acceleration of Financial Valuations

A High Performance Solution



- ❑ A Collaboration between Pricing Partners and ActiveEon
- ❑ Price-it® Excel Accelerated by *ProActive Parallel Suite®*
 - A Global Solution: fully integrated with the same functionalities and interface as Price-it Excel while increasing its computing power
 - High Quality Service: from both companies


Some Technical Facts



❑ Price-It®

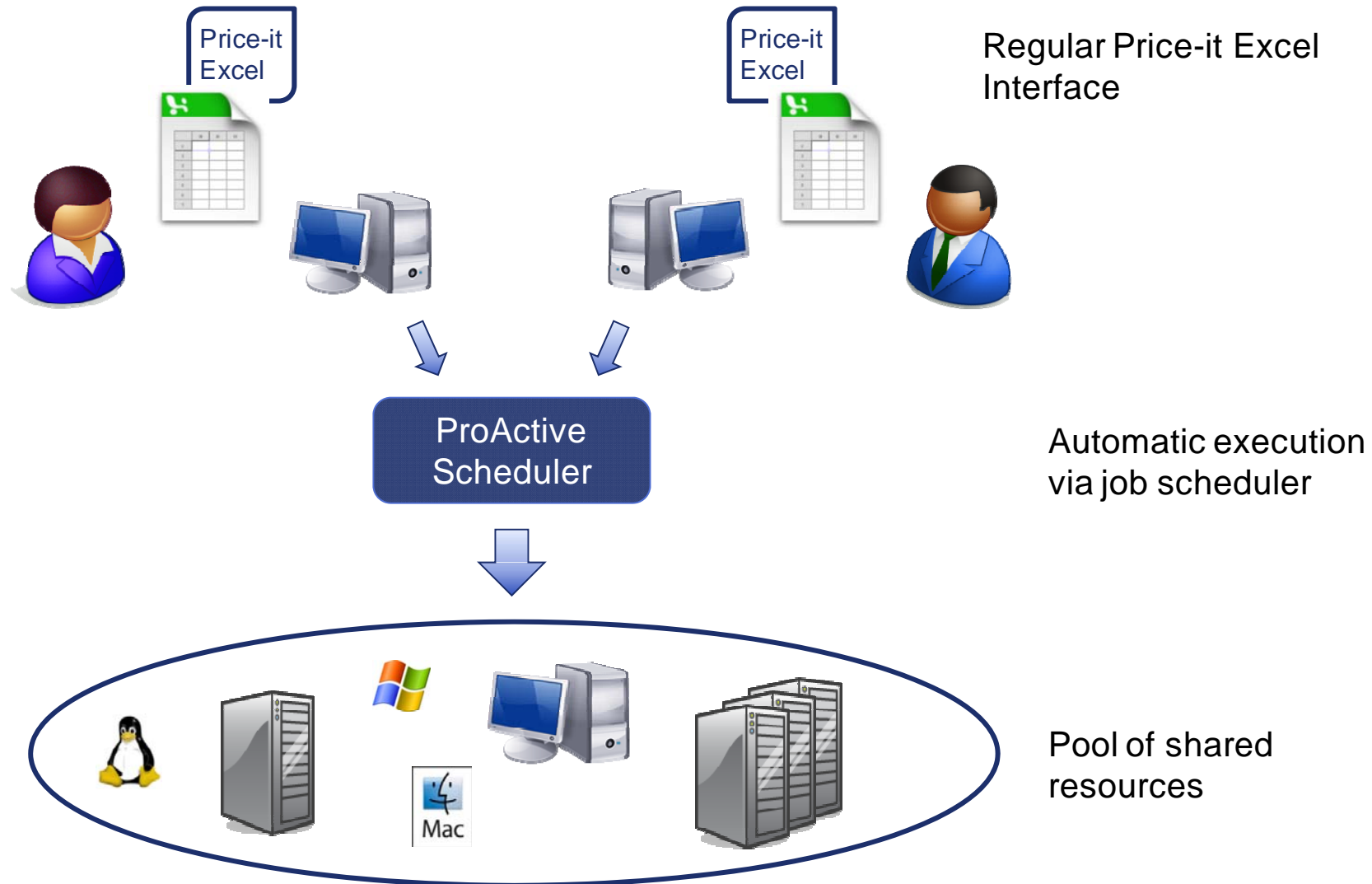
- C++ library developed by Pricing Partners
- Pricing solution dedicated to highly complex financial derivatives

❑ Specification and Constraints

- Accelerate Price-It® Excel product
 - Built on Price-It® library, this product integrates an interface with Excel for input data management and results display 
- Focus on highly parallelizable Greek computation
- Operating system: Windows

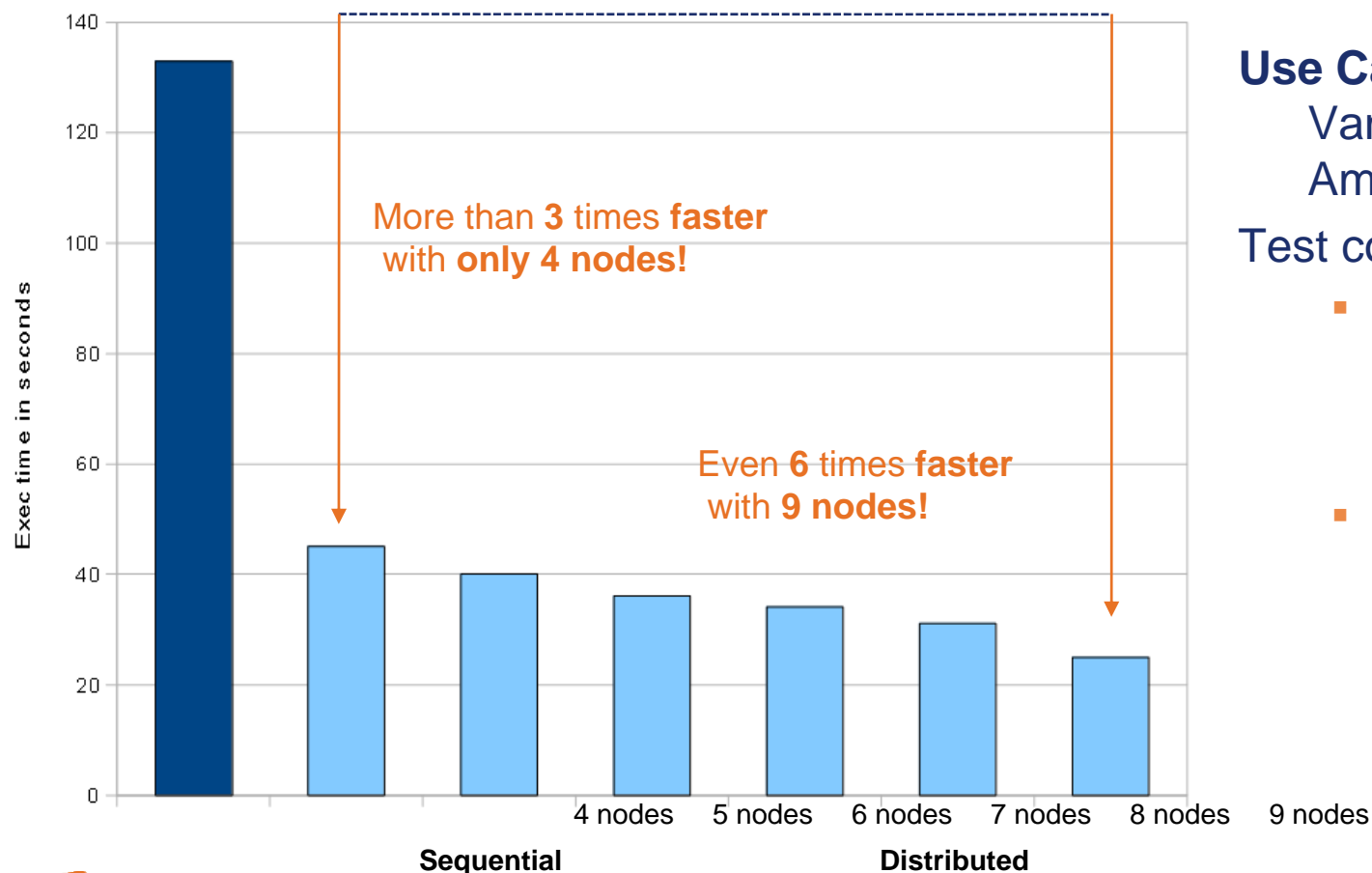
How Does it Work?

Price-it Computing Distribution



Accelerated Price-it Performances

- ❑ **Increased Productivity:** Reduces Price-it Execution Time by 6 or more!



Use Case: Bermuda
Vanilla, Model
American MC

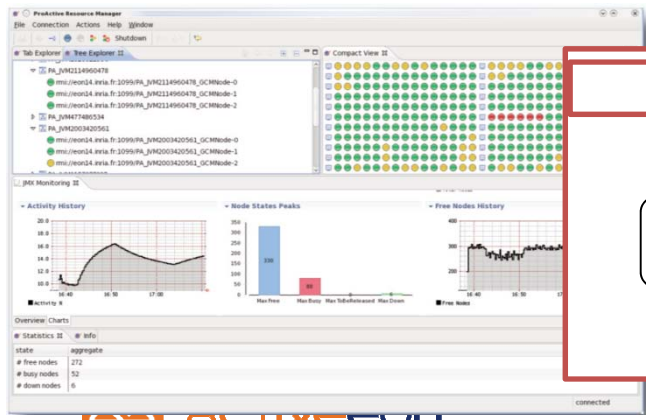
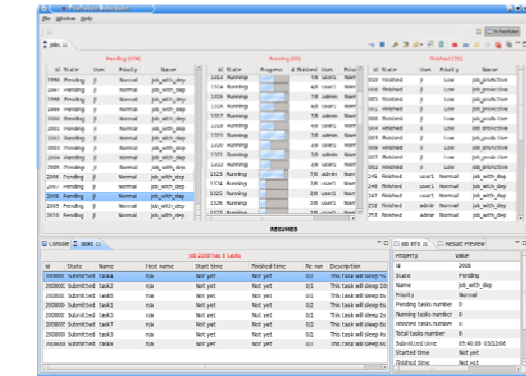
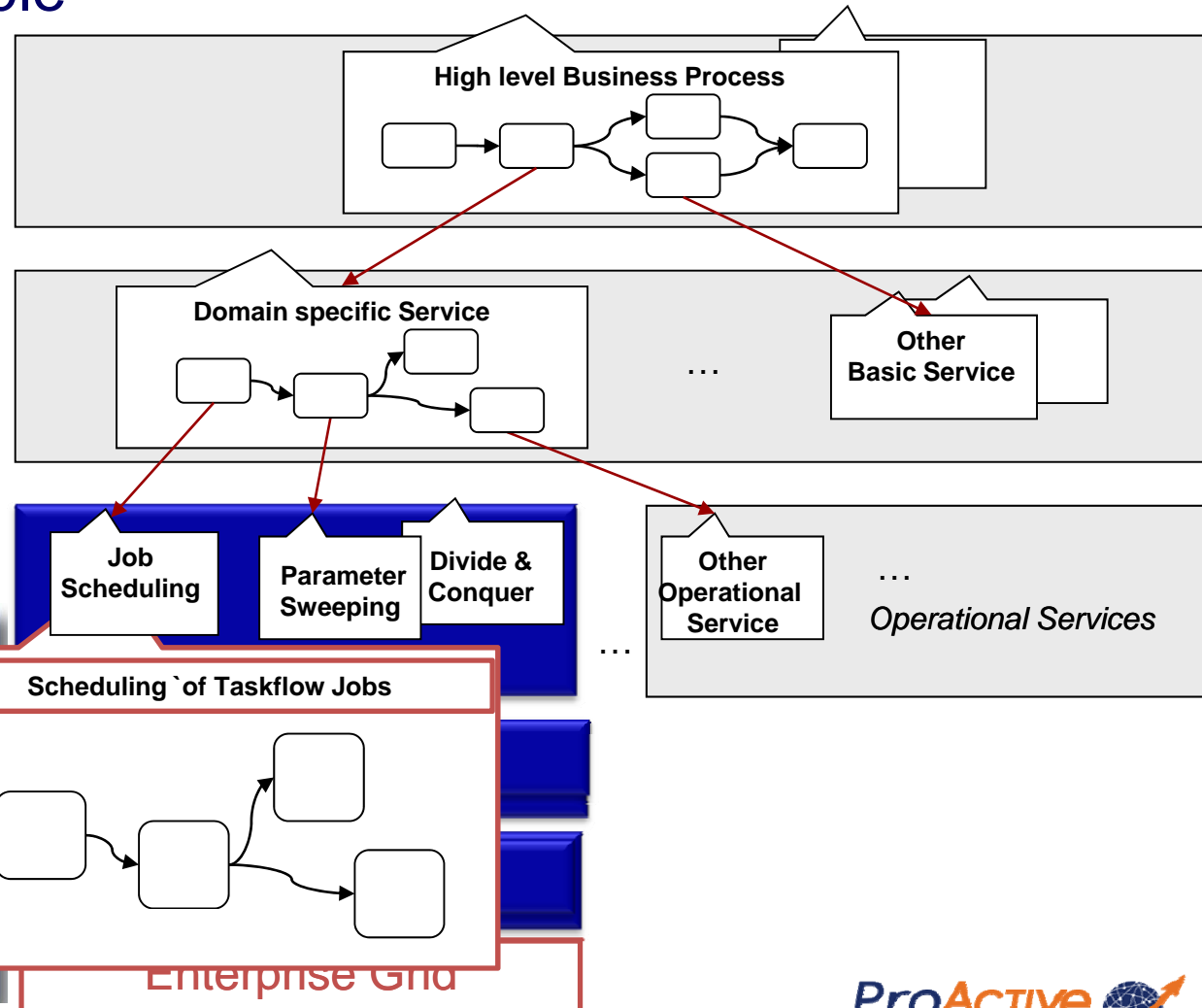
Test conditions:

- One computation is split in 130 tasks that are distributed
- Each task uses 300ko

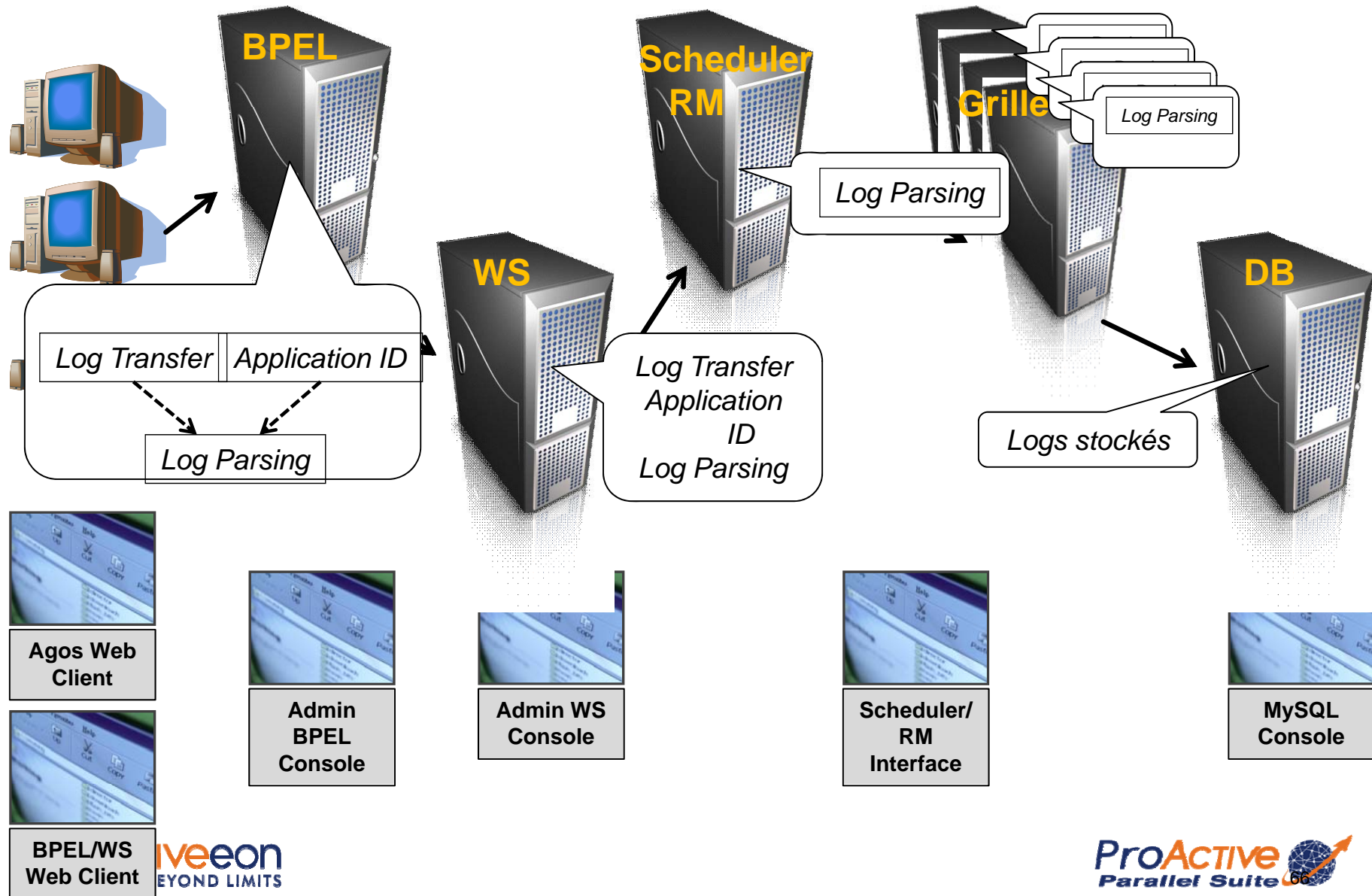
UC 3: SOA Analysis of Web Server Logs

Parallel Services

- Separation: BPEL – Parallel Serv. – Task Flow
- Standards et Portable
- Flexibility



Cas d'études AMADEUS : Démonstration



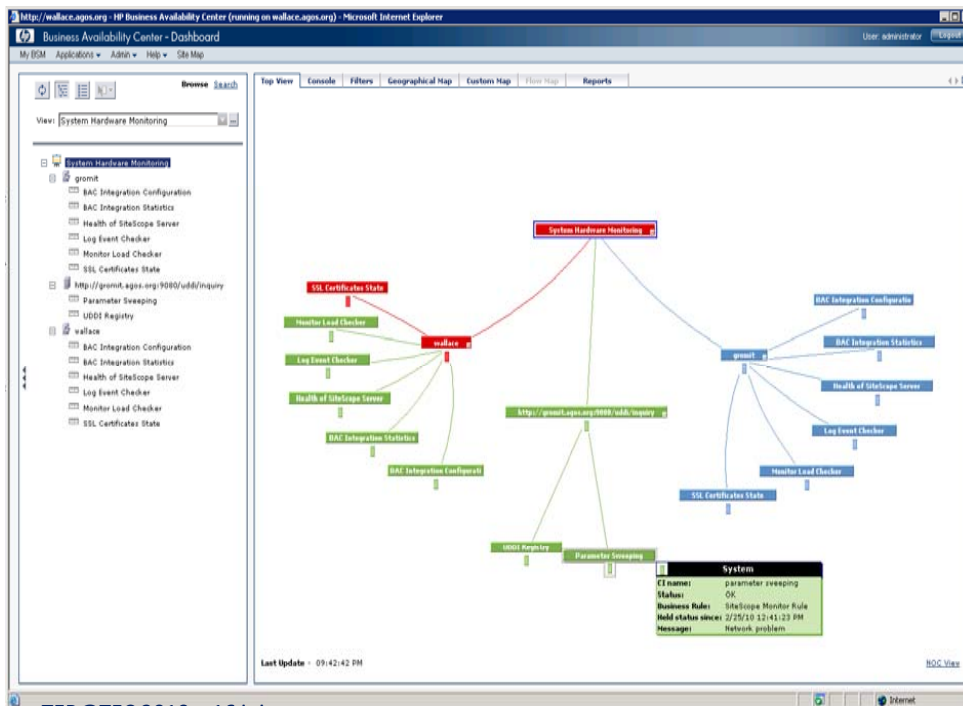
Grid Monitoring integration



AGOS Platform Management

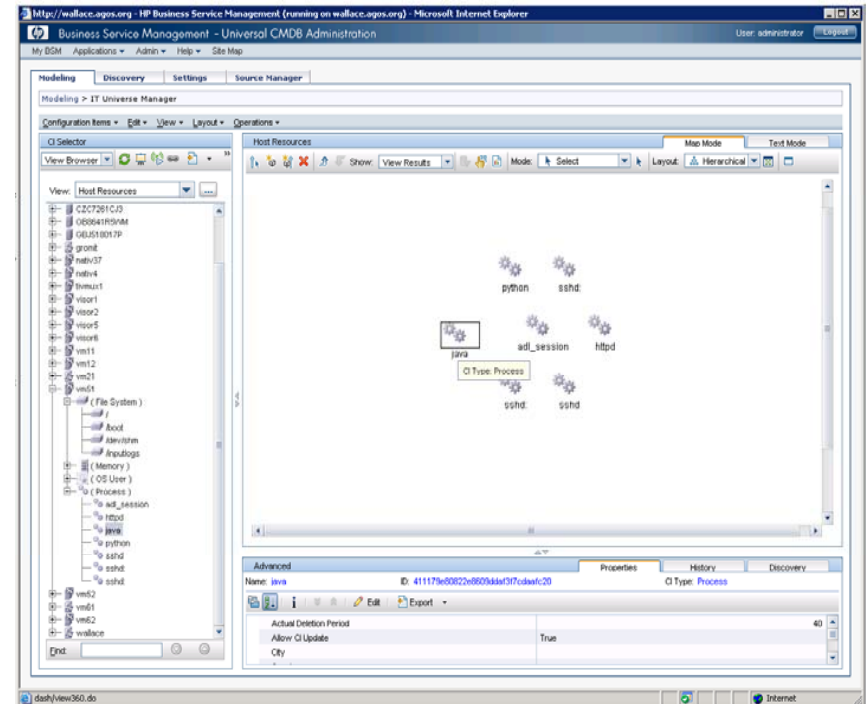
HP- Business Availability Center (HP-BAC)

- Monitoring of the entire platform
- Cover all layers in the scope
- Provide monitoring dashboard and reports



Tasks scheduler & Resources manager

- Integration with grid components
- Grid insights through indicator collection and running jobs on



AGOS: Grid Architecture for SOA

Building a Platform for Agile SOA with Grid

□ AGOS Solutions



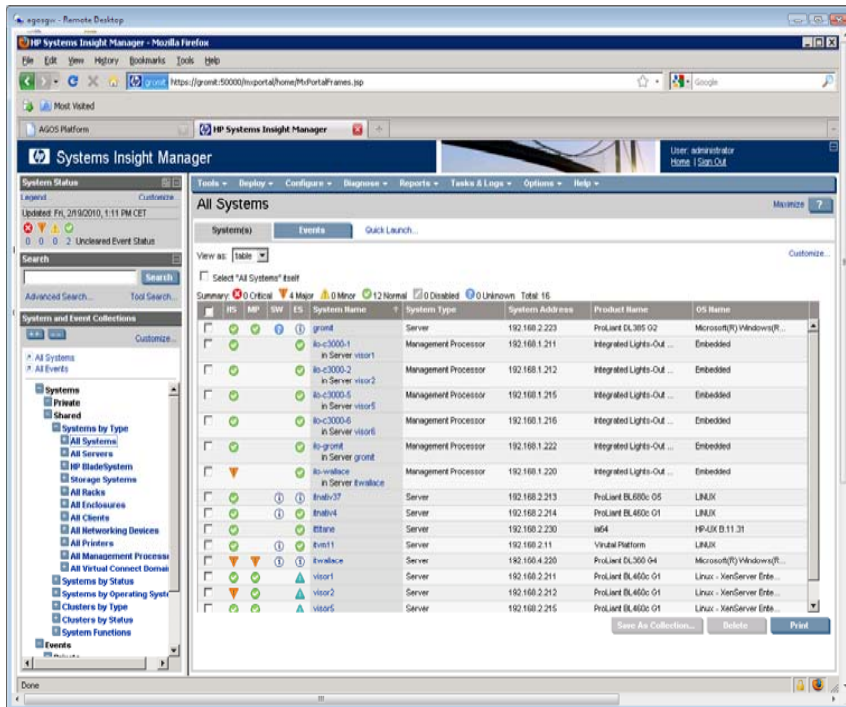
In Open Source with Professional Support



AGOS Infrastructure Management

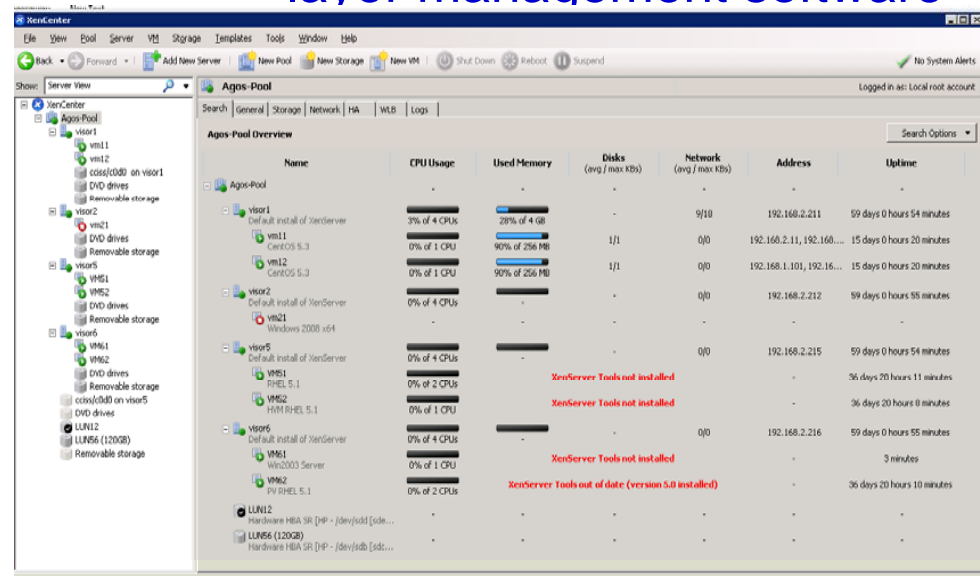
HP Systems Insight Manager (HP-SIM)

- Monitoring of entire infrastructure
- Communicates with upper layer management software (HP BAC)

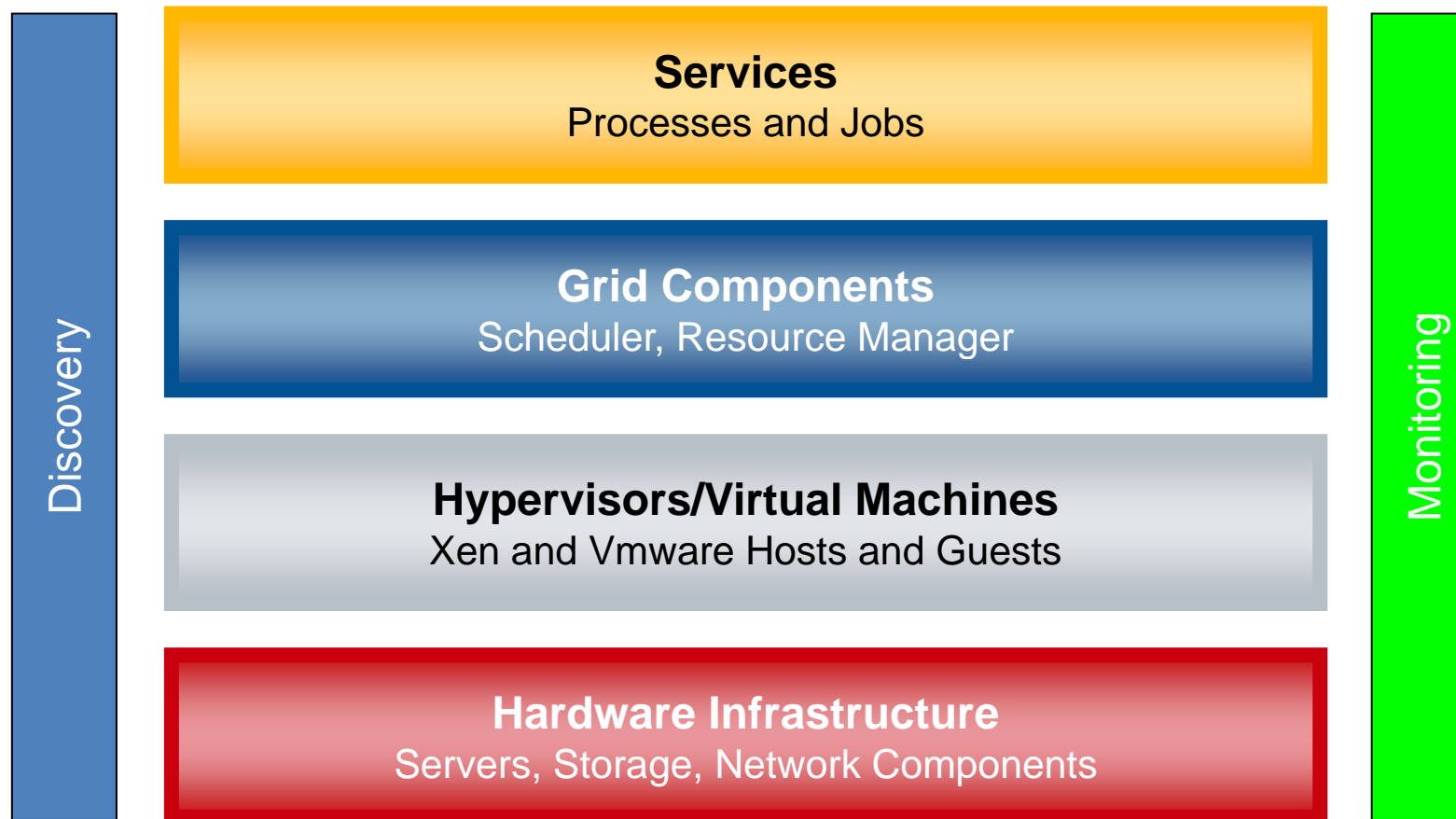


Citrix XenCenter

- Hypervisor and VM management
- Communicates with upper layer management software

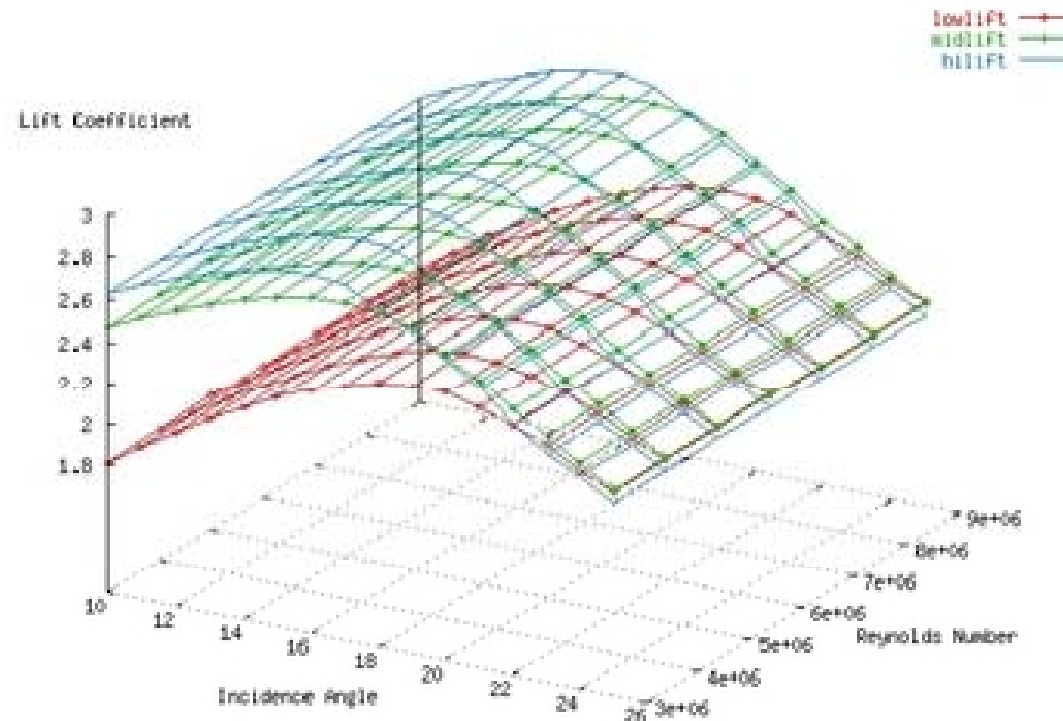
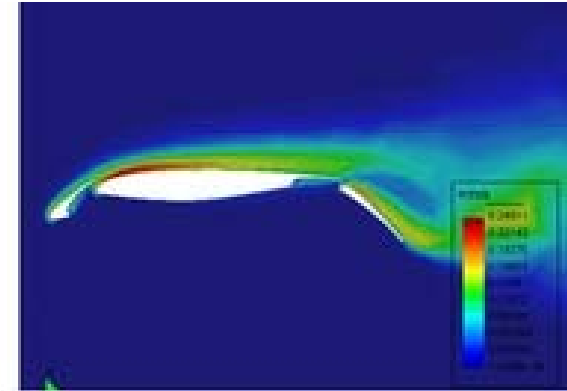
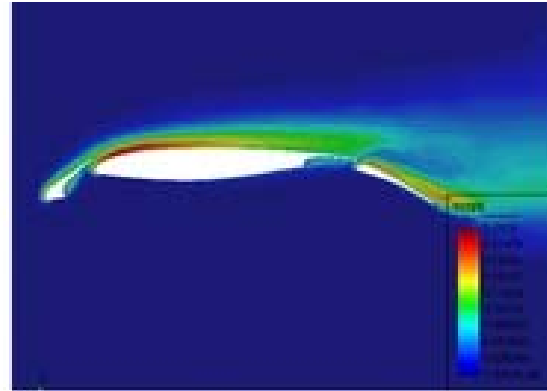
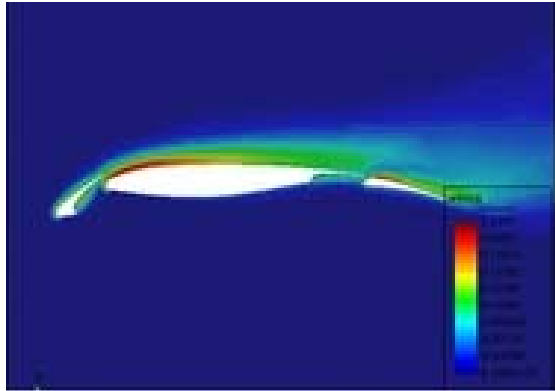


AGOS and HP Management tools Integration

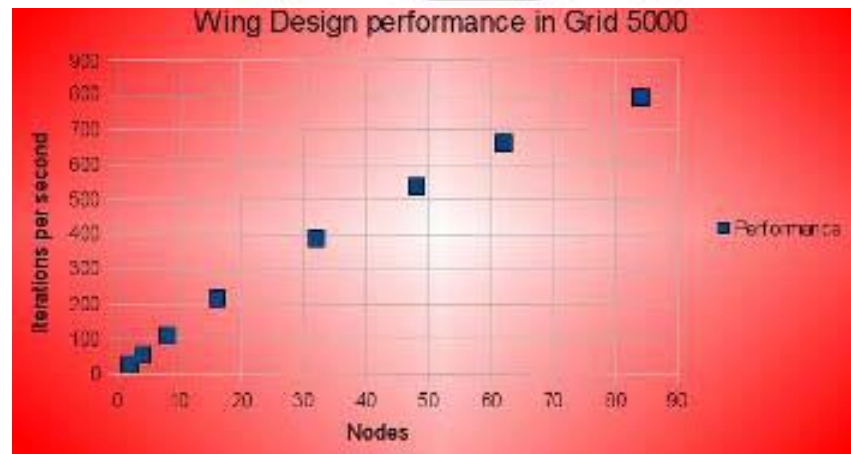
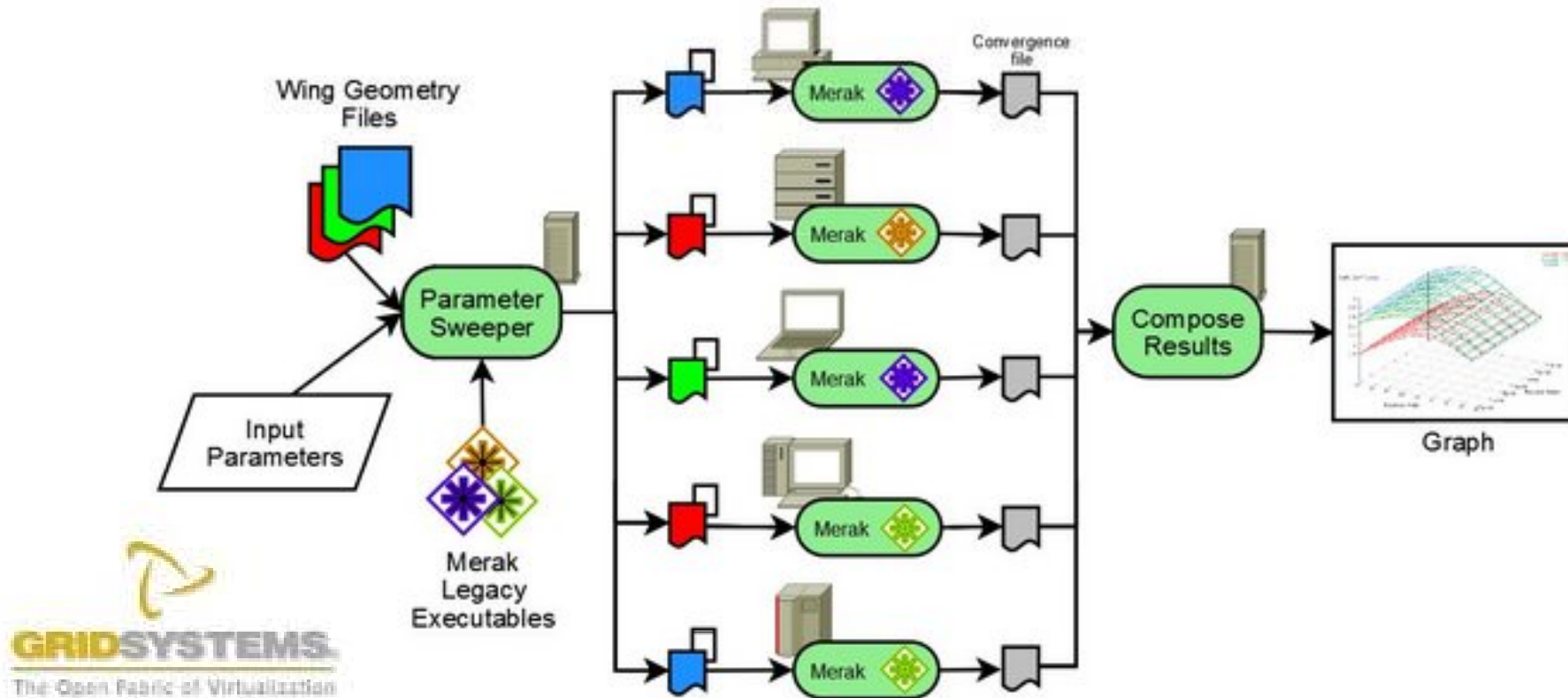


4. Use Case: Wings Optimization

Wing performance (GridSystems, Spain)



Wing performance (GridSystems, Spain)



(2) ASP: Asynchronous Sequential Processes

$$\frac{(a, \sigma) \rightarrow_S (a', \sigma')}{\alpha[a; \sigma; \iota; F; R; f] \parallel P \longrightarrow \alpha[a'; \sigma'; \iota; F; R; f] \parallel P} \text{ (LOCAL)}$$

Local

$$\frac{\begin{array}{l} \gamma \text{ fresh activity} \quad \iota' \notin \text{dom}(\sigma) \quad \sigma' = \{\iota' \mapsto AO(\gamma)\} :: \sigma \\ \sigma_\gamma = \text{copy}(\iota'', \sigma) \quad \text{Service} = (\text{if } m_j = \emptyset \text{ then } \text{FifoService} \text{ else } \iota''.m_j()) \end{array}}{\alpha[\mathcal{R}[\text{Active}(\iota'', m_j)]; \sigma; \iota; F; R; f] \parallel P \longrightarrow \alpha[\mathcal{R}[\iota']; \sigma'; \iota; F; R; f] \parallel \gamma[\text{Service}; \sigma_\gamma; \iota''; \emptyset; \emptyset; \emptyset] \parallel P} \text{ (NEWACT)}$$

Creating an Activity

$$\frac{\begin{array}{l} \sigma_\alpha(\iota) = AO(\beta) \quad \iota'' \notin \text{dom}(\sigma_\beta) \quad f_i^{\alpha \rightarrow \beta} \text{ new future} \quad \iota_f \notin \text{dom}(\sigma_\alpha) \\ \sigma'_\beta = \text{Copy\&Merge}(\sigma_\alpha, \iota'; \sigma_\beta, \iota'') \quad \sigma'_\alpha = \{\iota_f \mapsto \text{fut}(f_i^{\alpha \rightarrow \beta})\} :: \sigma_\alpha \end{array}}{\alpha[\mathcal{R}[\iota.m_j(\iota')]; \sigma_\alpha; \iota_\alpha; F_\alpha; R_\alpha; f_\alpha] \parallel \beta[a_\beta; \sigma_\beta; \iota_\beta; F_\beta; R_\beta; f_\beta] \parallel P \longrightarrow \alpha[\mathcal{R}[\iota_f]; \sigma'_\alpha; \iota_\alpha; F_\alpha; R_\alpha; f_\alpha] \parallel \beta[a_\beta; \sigma'_\beta; \iota_\beta; F_\beta; R_\beta; f_\beta] :: [m_j; \iota''; f_i^{\alpha \rightarrow \beta}]; f_\beta] \parallel P} \text{ (REQUEST)}$$

Sending a Request

$$\frac{R = R' :: [m_j; \iota_r; f'] :: R'' \quad m_j \in M \quad \forall m \in M, m \notin R'}{\alpha[\mathcal{R}[\text{Serve}(M)]; \sigma; \iota; F; R; f] \parallel P \longrightarrow \alpha[\iota.m_j(\iota_r) \uparrow f, \mathcal{R}[\Box]; \sigma; \iota; F; R' :: R''; f'] \parallel P} \text{ (SERVE)}$$

Service

$$\frac{\iota' \notin \text{dom}(\sigma) \quad F' = F :: \{f \mapsto \iota'\} \quad \sigma' = \text{Copy\&Merge}(\sigma, \iota; \sigma, \iota')}{\alpha[\iota \uparrow (f', a); \sigma; \iota; F; R; f] \parallel P \longrightarrow \alpha[a; \sigma'; \iota; F'; R; f'] \parallel P} \text{ (ENDSERVICE)}$$

□

$$\frac{\begin{array}{l} \sigma_\alpha(\iota) = \text{fut}(f_i^{\gamma \rightarrow \beta}) \quad F_\beta(f_i^{\gamma \rightarrow \beta}) = \iota_f \quad \sigma'_\alpha = \text{Copy\&Merge}(\sigma_\beta, \iota_f; \sigma_\alpha, \iota) \end{array}}{\alpha[a_\alpha; \sigma_\alpha; \iota_\alpha; F_\alpha; R_\alpha; f_\alpha] \parallel \beta[a_\beta; \sigma_\beta; \iota_\beta; F_\beta; R_\beta; f_\beta] \parallel P \longrightarrow \alpha[a_\alpha; \sigma'_\alpha; \iota_\alpha; F_\alpha; R_\alpha; f_\alpha] \parallel \beta[a_\beta; \sigma_\beta; \iota_\beta; F_\beta; R_\beta; f_\beta] \parallel P} \text{ (REPLY)}$$

Sending a Reply

Key Point

- ❑ **“MPI and programming languages from the 60’s will not make it”**
- ❑ **Jack Dongarra, 2/13/2009,**
- ❑ **Wake Forest University talk**

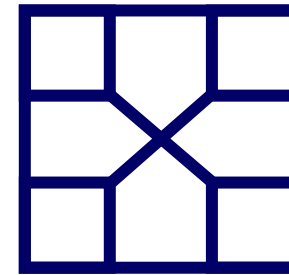
NAS Parallel Benchmarks

- Experimented on 3D ElectroMagnetism, and Nasa Benchmarks
- ❑ **Designed by NASA to evaluate benefits of high performance systems**
- ❑ **Strongly based on CFD**
- ❑ **5 benchmarks (kernels) to test different aspects of a system**
- ❑ **2 categories or focus variations:**
 - **communication intensive and computation intensive**

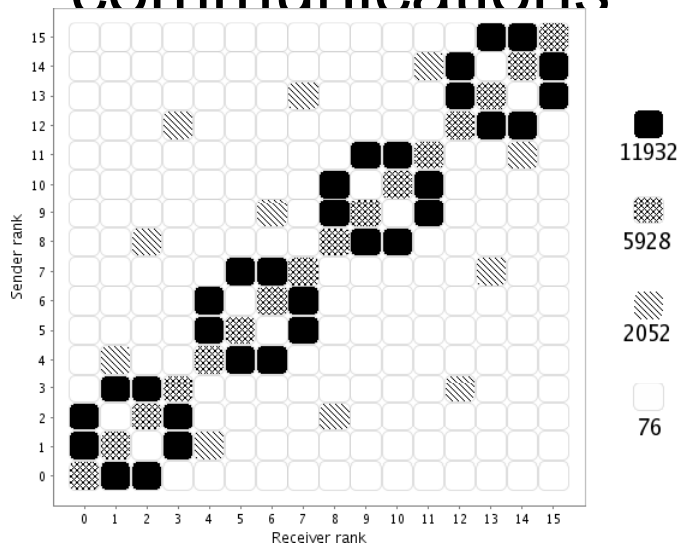


Communication Intensive CG Kernel (Conjugate Gradient)

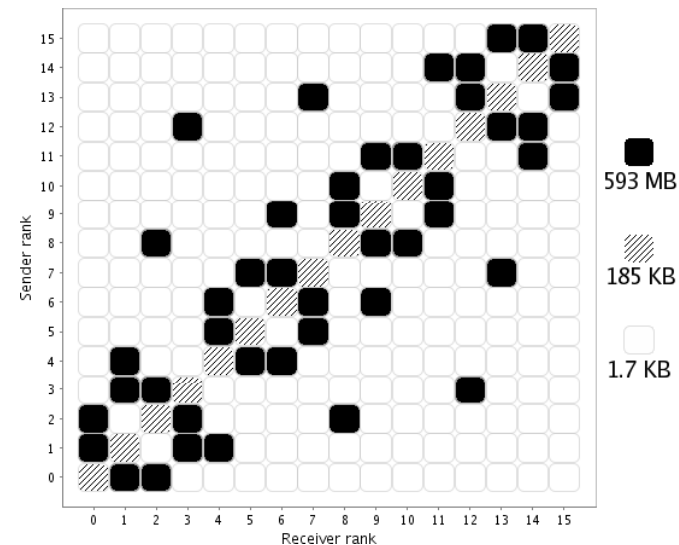
- ❑ Floating point operations
- ❑ Eigen value computation
- ❑ High number of unstructured communications



- 12000 calls/node
- 570 MB sent/node
- 1 min 32
- 65 % comms/WT

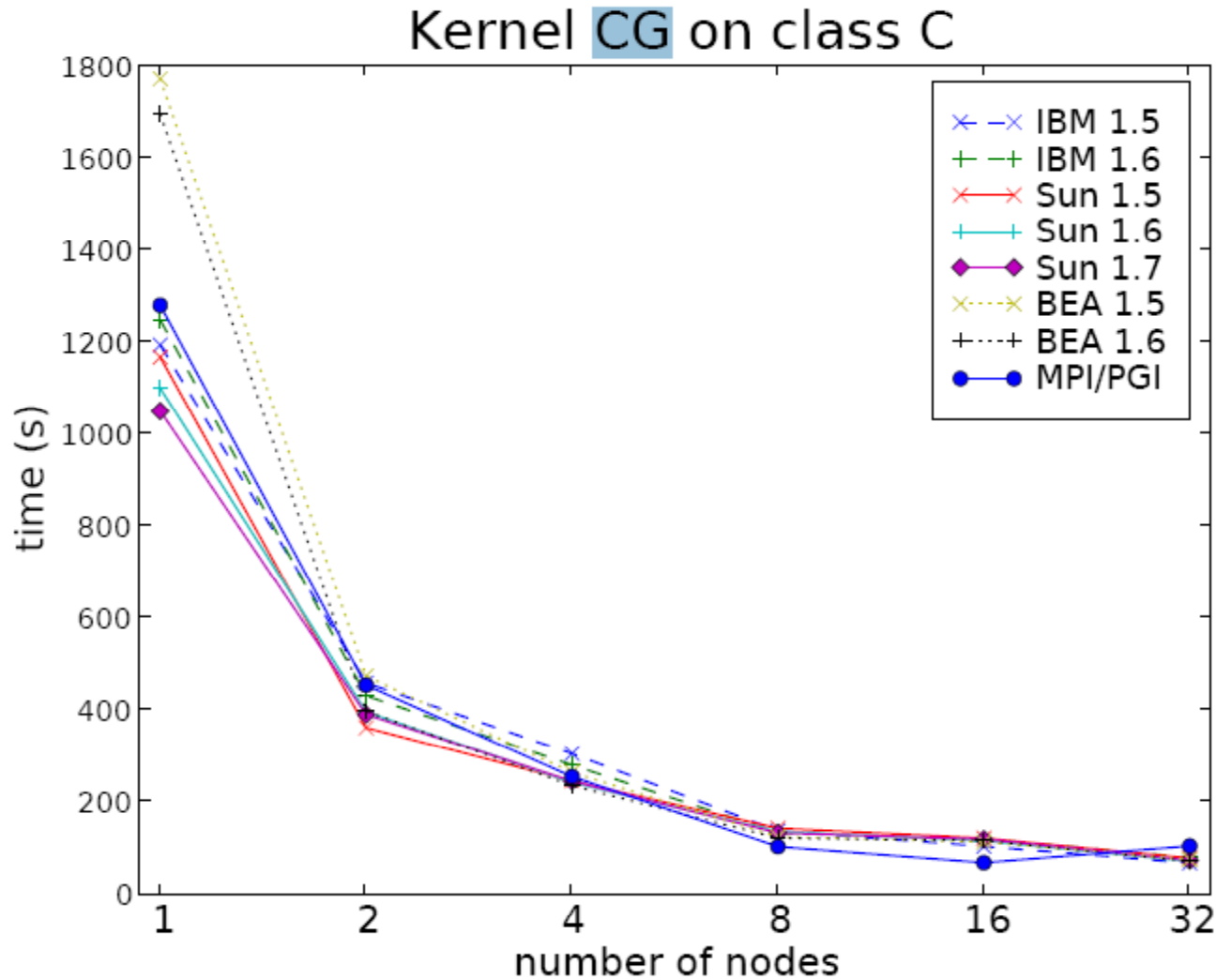


Message density distribution



Data density distribution

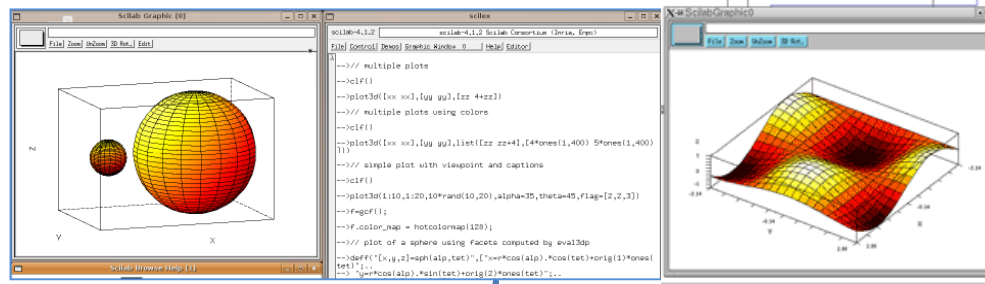
Communication Intensive CG Kernel (Conjugate Gradient)



← Comparable Performances

3. Parallel Scilab Simulations

Seamless Parallel & Distributed Scilab



Static Policy

LSF



Dedicated resources

**Timing Policy
12/24**

Desktops



Desktops

**Dynamic
Workload Policy**

EC2



Amazon EC2



Forum TER@TEC 2010 – 16 juin



Interface ProActive ⇔ Scilab

The screenshot displays the ProActive Scheduler interface, which is divided into several panes. The top pane shows the 'Jobs' status, categorized into Pending (8), Running (10), and Finished (7). The bottom pane shows the 'Tasks' status for a specific job (Job 65), listing task IDs, states, names, start times, finished times, run time limits, and descriptions. The rightmost pane shows the 'Jobs info' for Job 65, providing details such as task counts, submission time, and execution duration.

Console Output:

```
-->PAconnect('rmi://shainese.inria.fr:6608');
Connection successful to rmi://shainese.inria.fr:6608
-->res1 = PAolve('cosh', list(1,2,3,4,5,6,7,8,9,10));
-->res1
res1 =
    res1(1)
    1.5430806
    res1(2)
    3.7621957
    res1(3)
    10.067662
    res1(4)
    27.308233
    res1(5)
    74.209949
    res1(6)
    201.71564
    res1(7)
    548.31704
    res1(8)
    1490.4792
    res1(9)
    4051.542
```

Jobs Status:

Pending (8)				Running (10)				Finished (7)				
Id	Priority	Name	Description	Id	Task	Priority	Name	Description	Id	Priority	Name	Description
78	Normal	job_2_tasks	2 tasks with variable durations	68	7/8	Normal	job_8_tasks	Simple test of 8 tasks with	61	Normal	job_8_tasks	Simple test of 8 tasks with variat
79	Normal	job_2_tasks	2 tasks with variable durations	69	7/8	Normal	job_8_tasks	Simple test of 8 tasks with	62	Normal	job_8_tasks	Simple test of 8 tasks with variat
80	Normal	job_2_tasks	2 tasks with variable durations	70	6/8	Normal	job_8_tasks	Simple test of 8 tasks with	63	Normal	job_8_tasks	Simple test of 8 tasks with variat
81	Normal	job_2_tasks	2 tasks with variable durations	71	0/1	Normal	job_PI	Calcul de Pi, methode de l	64	Normal	job_8_tasks	Simple test of 8 tasks with variat
82	Normal	job_2_tasks	2 tasks with variable durations	72	0/1	Normal	job_PI	Calcul de Pi, methode de l	65	Normal	job_8_tasks	Simple test of 8 tasks with variat
83	Normal	job_2_tasks	2 tasks with variable durations	73	0/1	Normal	job_PI	Calcul de Pi, methode de l	66	Normal	job_8_tasks	Simple test of 8 tasks with variat
84	Normal	job_2_tasks	2 tasks with variable durations	74	0/1	Normal	job_PI	Calcul de Pi, methode de l	67	Normal	job_8_tasks	Simple test of 8 tasks with variat
85	Normal	job_2_tasks	2 tasks with variable durations	75	0/1	Normal	job_PI	Calcul de Pi, methode de l				
				76	1/2	Normal	job_2_tasks	2 tasks with variable durat				
				77	0/2	Normal	job_2_tasks	2 tasks with variable durat				

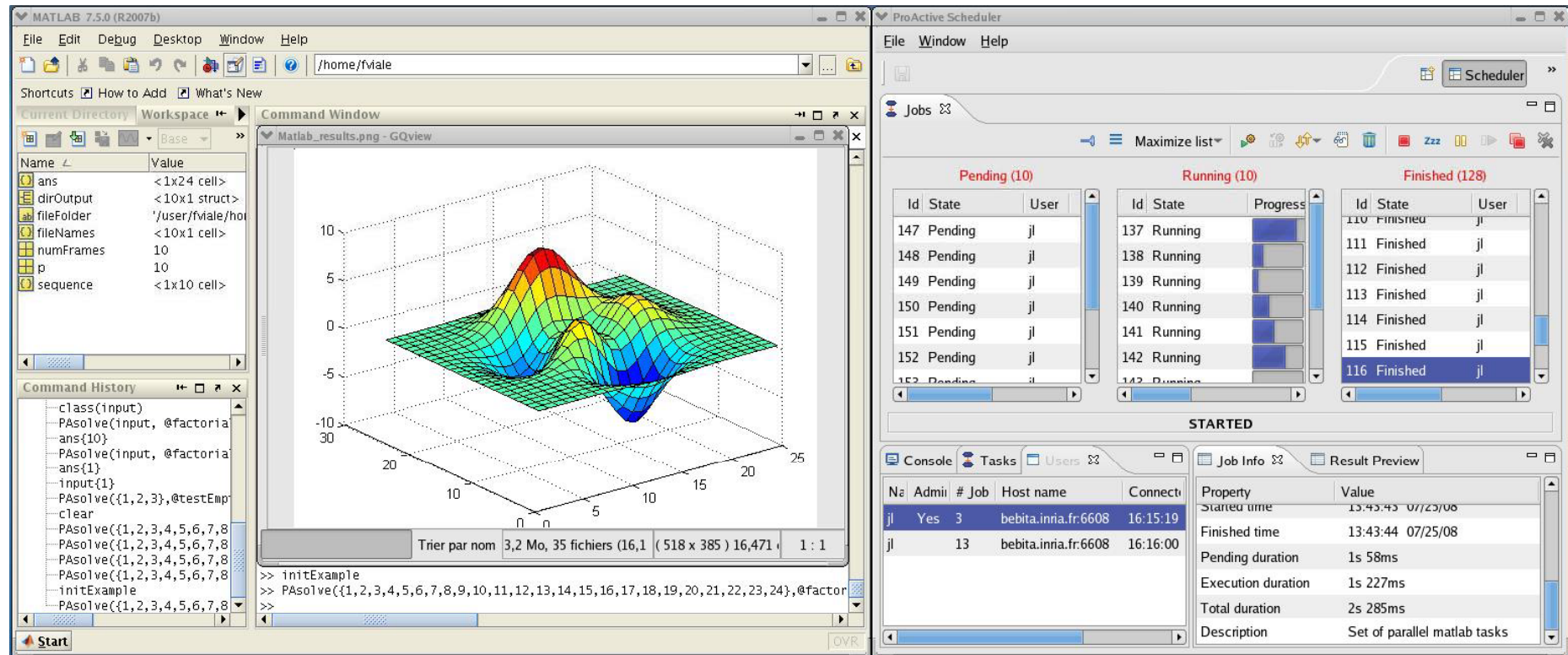
Tasks Status (Job 65 has 8 tasks):

Id	State	Name	Start time	finished time	Run time limit	Renunnable	Description
65001	Finished	task6	08:55:11 07/05/07	08:55:16 07/05/07			
65002	Finished	task5	08:55:13 07/05/07	08:55:21 07/05/07			
65003	Finished	task4	08:55:14 07/05/07	08:55:20 07/05/07			
65004	Finished	task2	08:55:14 07/05/07	08:55:21 07/05/07			
65005	Finished	task8	08:55:15 07/05/07	08:55:35 07/05/07			
65006	Finished	task1	08:55:15 07/05/07	08:55:23 07/05/07			
65007	Finished	task3	08:55:16 07/05/07	08:55:24 07/05/07			
65008	Finished	task7	08:55:17 07/05/07	08:55:22 07/05/07			

Jobs info (Job 65):

Property	Value
Id	65
Name	job_8_tasks
Priority	Normal
Pending tasks number	0
Running tasks number	0
Finished tasks number	8
Total tasks number	8
Submitted time	08:54:55 07/05/07
Started time	08:55:11 07/05/07
Finished time	08:55:35 07/05/07
Pending duration	16s 25ms
Execution duration	24s 622ms

Interface ProActive ⇔ Matlab



Interface ProActive \Leftrightarrow Scilab / Matlab

□ Primitives de haut niveau Scilab/Matlab pour distribuer les calculs :

- `f = PAeval(@foo,p1)`
// evaluate `f(p1)` en asynchrone
// `f` est un futur
- `L = PAsolve(@foo, {p1,p2, ... pk})`
// evaluate `foo(p1), foo(p2) .. foo(pk)` en asynchrone
// `L = f1,...fk` est une liste de futurs

Synchronization des calculs

□ Primitives d'attente :

- $r = \text{PAwaitForAny}(L)$
// bloque et attend jusqu'à l'obtention du
// premier des fk et retourne le resultat r
- $lr = \text{PAgetAny}(L)$
// non-blocant, retourne une liste des resultats
// deja obtenus

4. Use Case: OMD2

OMD2

Interfaces Open Sources pour faciliter les Optimisations Multi-Disciplinaires Distribuées



OMD2 :

Interfaces Open Sources pour faciliter les Optimisations Multi-Disciplinaires Distribuées

OMD2
OPTIMISATION
MULTI
DISCIPLINAIRE
DISTRIBUEE



Laboratoire Roberval
Unité de recherche en mécanique



Ecole Nationale
Supérieure des Mines
SAINT-ETIENNE

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



INRIA



digiteo
Research in information sciences and technologies

SIREHNA
a DCNS company



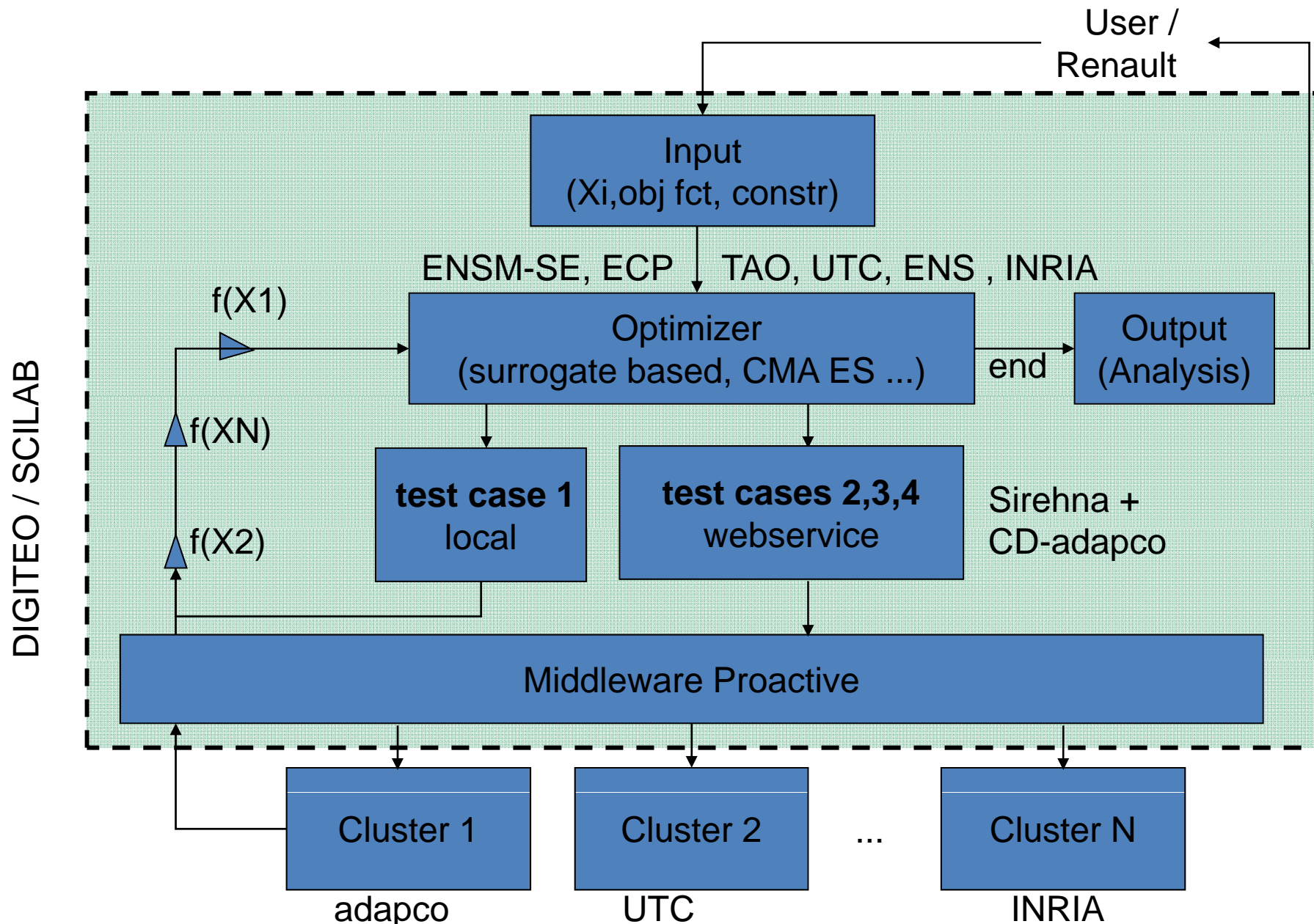
Activeeon
SCALE BEYOND LIMITS



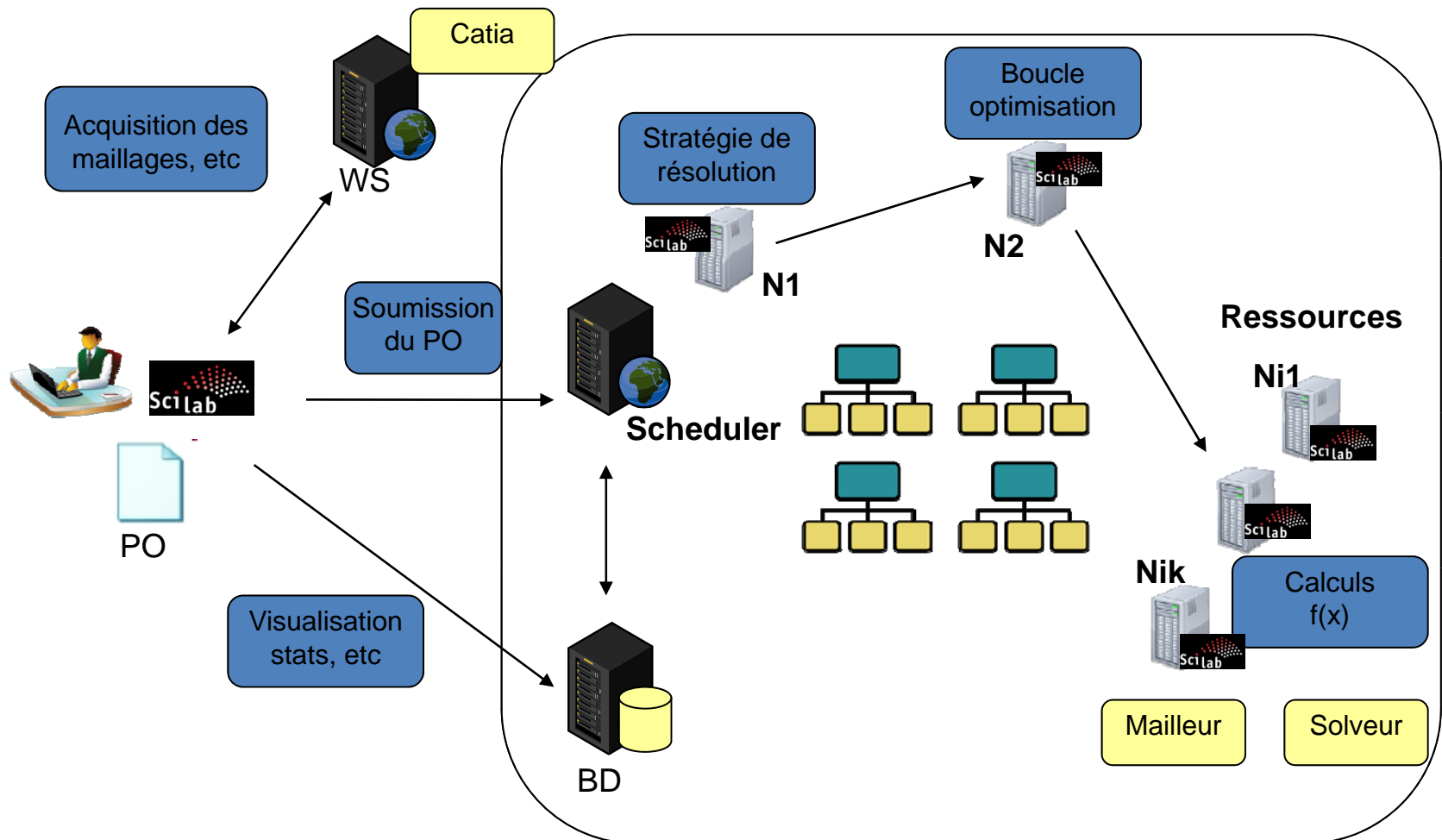
ProActive
Parallel Suite

OW2
Consortium

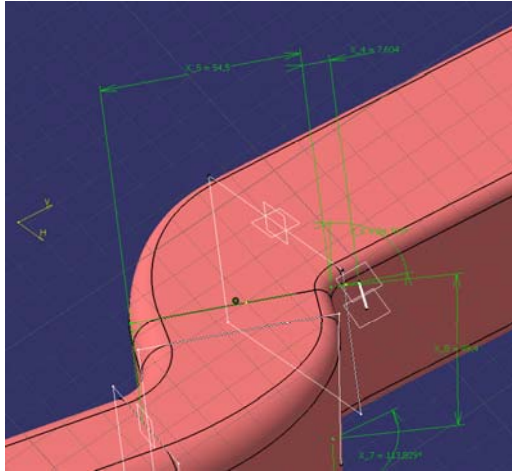
Plateforme OMD2, vue générale :



Distribution des calculs

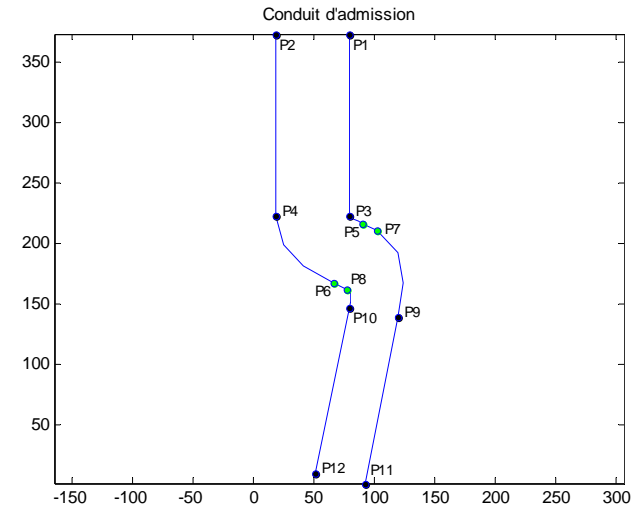


Les cas tests



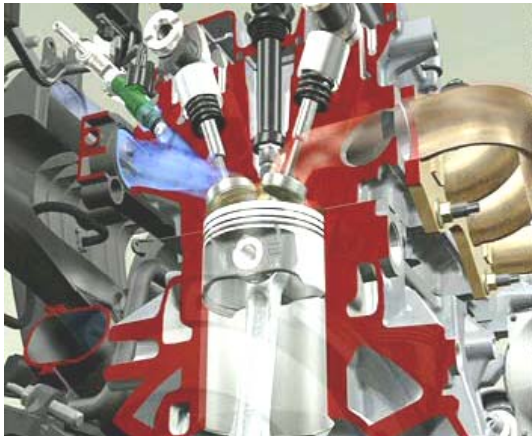
Conduit de climatisation 3D

10min
CPU



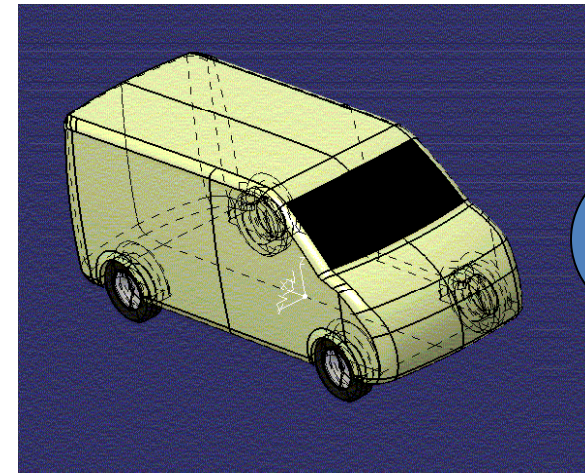
Conduit de climatisation 2D

<1min
CPU



Culasse multi-disciplinaire

100h
CPU



Aéro externe

1000h
CPU

Interface ProActive ↔ Scilab

Console

```
-->PAconnect('rmi://shainese.inria.fr:6608');
Connection successful to rmi://shainese.inria.fr:6608
-->res1 = PAolve('cosh', list(1,2,3,4,5,6,7,8,9,10));
-->res1
res1 =
    res1(1)
    1.5430806
    res1(2)
    3.7621957
    res1(3)
    10.067662
    res1(4)
    27.308233
    res1(5)
    74.209949
    res1(6)
    201.71564
    res1(7)
    548.31704
    res1(8)
    1490.4792
    res1(9)
    4051.542
```

Scheduler

Pending (8)

Id	Priority	Name	Description
78	Normal	job_2_tasks	2 tasks with variable durations
79	Normal	job_2_tasks	2 tasks with variable durations
80	Normal	job_2_tasks	2 tasks with variable durations
81	Normal	job_2_tasks	2 tasks with variable durations
82	Normal	job_2_tasks	2 tasks with variable durations
83	Normal	job_2_tasks	2 tasks with variable durations
84	Normal	job_2_tasks	2 tasks with variable durations
85	Normal	job_2_tasks	2 tasks with variable durations

Running (10)

Id	Task	Priority	Name	Description
68	7/8	Normal	job_8_tasks	Simple test of 8 tasks with
69	7/8	Normal	job_8_tasks	Simple test of 8 tasks with
70	6/8	Normal	job_8_tasks	Simple test of 8 tasks with
71	0/1	Normal	job_PI	Calcul de Pi, methode de l
72	0/1	Normal	job_PI	Calcul de Pi, methode de l
73	0/1	Normal	job_PI	Calcul de Pi, methode de l
74	0/1	Normal	job_PI	Calcul de Pi, methode de l
75	0/1	Normal	job_PI	Calcul de Pi, methode de l
76	1/2	Normal	job_2_tasks	2 tasks with variable durat
77	0/2	Normal	job_2_tasks	2 tasks with variable durat

Finished (7)

Id	Priority	Name	Description
61	Normal	job_8_tasks	Simple test of 8 tasks with variat
62	Normal	job_8_tasks	Simple test of 8 tasks with variat
63	Normal	job_8_tasks	Simple test of 8 tasks with variat
64	Normal	job_8_tasks	Simple test of 8 tasks with variat
65	Normal	job_8_tasks	Simple test of 8 tasks with variat
66	Normal	job_8_tasks	Simple test of 8 tasks with variat
67	Normal	job_8_tasks	Simple test of 8 tasks with variat

Jobs info

Property	Value
Id	65
Name	job_8_tasks
Priority	Normal
Pending tasks number	0
Running tasks number	0
Finished tasks number	8
Total tasks number	8
Submitted time	08:54:55 07/05/07
Started time	08:55:11 07/05/07
Finished time	08:55:35 07/05/07
Pending duration	16s 25ms
Execution duration	24s 622ms

Tasks

Job 65 has 8 tasks

Id	State	Name	Start time	finished time	Run time limit	Renunnable	Description
65001	Finished	task6	08:55:11 07/05/07	08:55:16 07/05/07			
65002	Finished	task5	08:55:13 07/05/07	08:55:21 07/05/07			
65003	Finished	task4	08:55:14 07/05/07	08:55:20 07/05/07			
65004	Finished	task2	08:55:14 07/05/07	08:55:21 07/05/07			
65005	Finished	task8	08:55:15 07/05/07	08:55:35 07/05/07			
65006	Finished	task1	08:55:15 07/05/07	08:55:23 07/05/07			
65007	Finished	task3	08:55:16 07/05/07	08:55:24 07/05/07			
65008	Finished	task7	08:55:17 07/05/07	08:55:22 07/05/07			

Exemple d'utilisation : Boucle d'Optimisation

```
futureList = {};  
PAconnect('//machine where the scheduler is running')  
do {  
    if size(futureList) < MAX_SENT {  
        x=opt.ask(); // asks a new vector  
        future = PAeval (@f, {x}); // asynchronously send f(x) to grid  
        futureList = append(futureList, future); }  
    else {  
        resultList = PAgetAny(futureList); // Get Available results  
        If resultList != void { // Updates the optimizer with new results  
            for i=1:size(resultList) {  
                [y,x] = resultList(i);  
                opt.tell(y,x);  
            }  
        } else Sleep (t)  
    }  
}  
while ~(ont.ston())
```

