

Federated Cloud Computing Environment for Malaria Fighting

Vilnius April-11-2011

Aurelio Rodriguez, Carlos Fernández, Ruben Díez, Hugo Gutierrez and
Álvaro Simón

plan
avanza2,»»



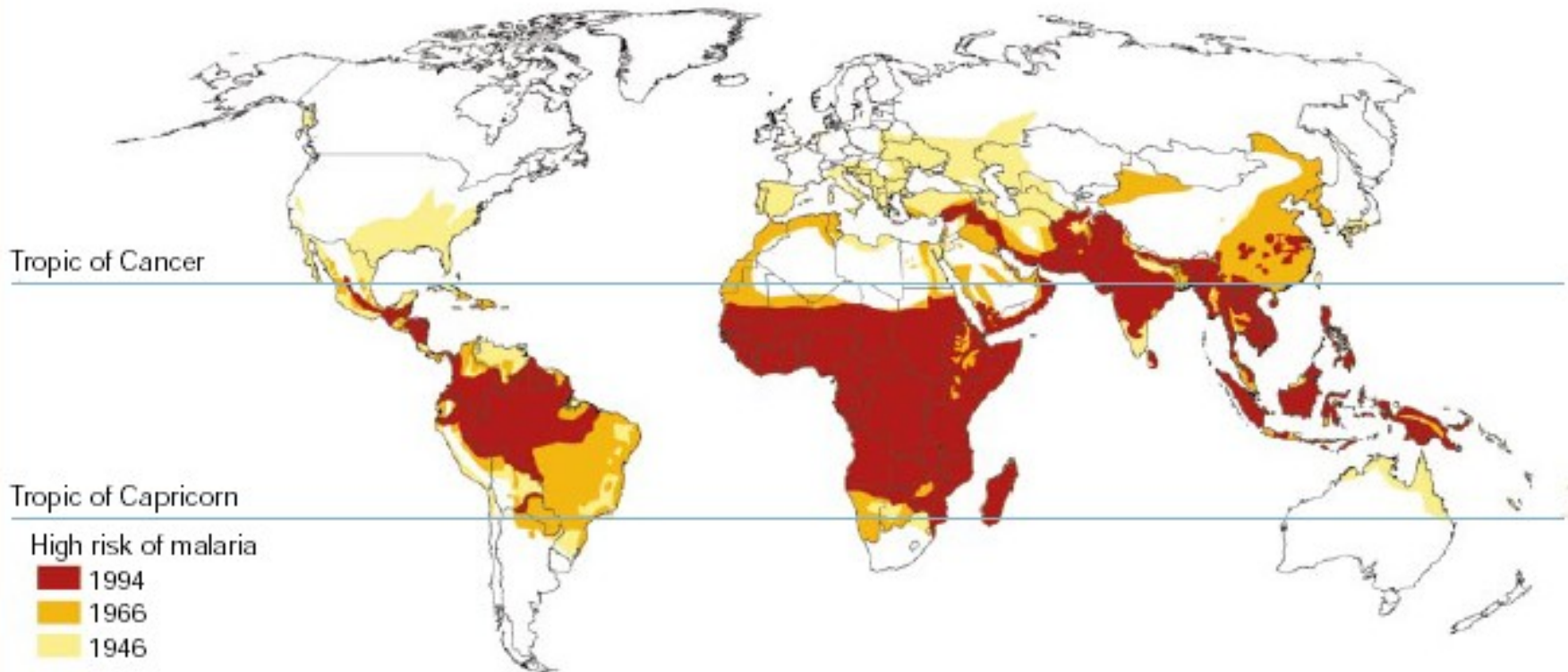
Proyecto parcialmente subvencionado por el subprograma Avanza I+D de la Acción Estratégica de Telecomunicaciones y Sociedad de la Información del Ministerio de Industria, Turismo y Comercio de España. Número de proyecto: TSI-020301-2009-30

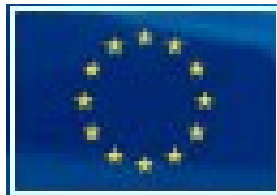
- **Introduction**
 - Motivation.
 - About Synergy.
 - About NUBA.
- **Computer-Aided Drug Design.**
 - Synergy Collaboration Pilots.
 - Chemical Database.
 - Database Preparation.
- **Federated Cloud for HPC.**
 - The issue.
 - Hardware resources.
 - OpenNebula.
 - Virtual Clusters.
 - Network Configuration.
 - OpenNebula Frontend.
 - Experiment Results.
- **Conclusions.**

INTRODUCTION

Motivation

- **Third world disease.**
- **500 million cases per year.**
- **1.5 – 3 million deaths per year (children bellow 5!).**
- **Number of cases constantly increasing.**
- **Several therapeutic tools but all of them generate resistances.**





Scientists Against Malaria

Virtual Organisation for Drug Discovery

DouglasConnect



CERF by Rescentris, Inc.



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

UNIVERSITY OF
Cincinnati

Drug Discovery Center



UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI FARMACIA



MONASH University



Fundación Pública Galega
de Medicina Xenómica



inte:ligand

Your partner for in-silico drug discovery.

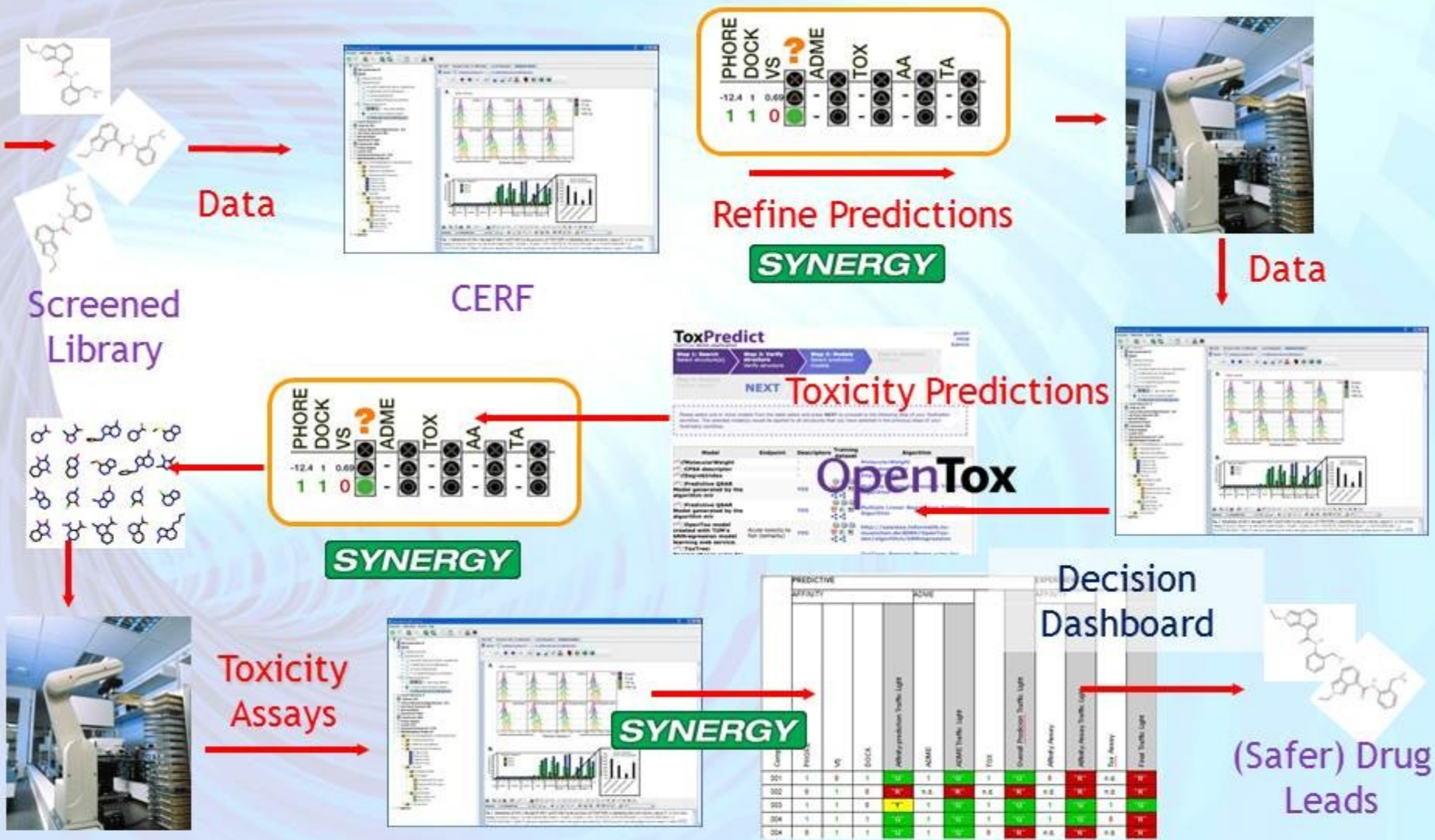
About NUBA

- **NUBA is a R+D+i project to develop a federated cloud computing platform (Infrastructure as Service).**
- **The new federated cloud platform will assist to deploy new Internet business services in an automated way.**
- **New services will be escalated dynamically based on business objectives and performance criterions.**
- **CESGA team is collaborating to deploy this new cloud infrastructure:**
 - **OpenNebula testbed and infrastructure coordination.**
 - **Cloud infrastructure monitoring and accounting.**
 - **E-IMRT use case (radiotherapy treatment planning on cloud).**



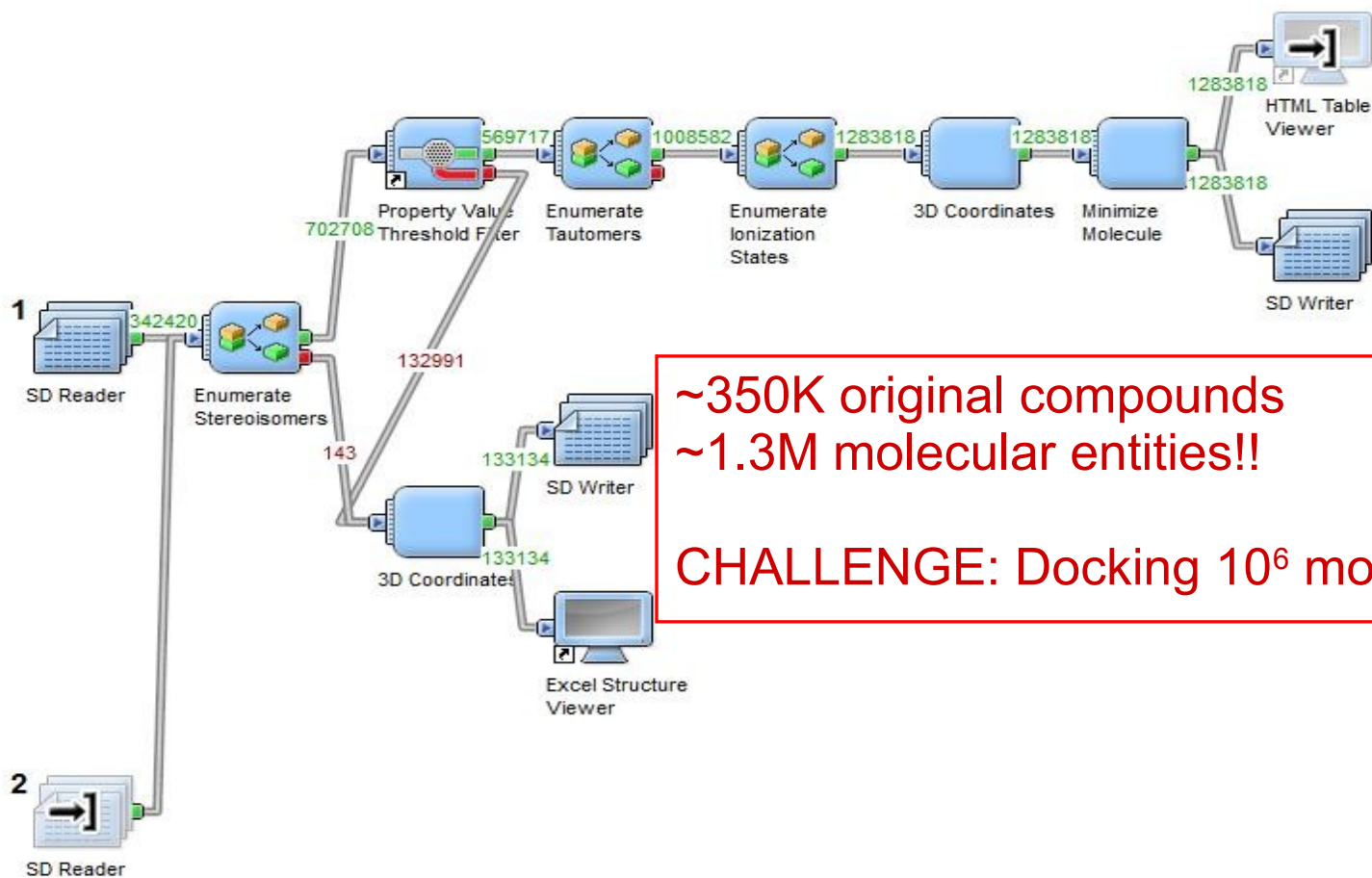
COMPUTER-AIDED DRUG DESIGN

Synergy Collaboration Pilots

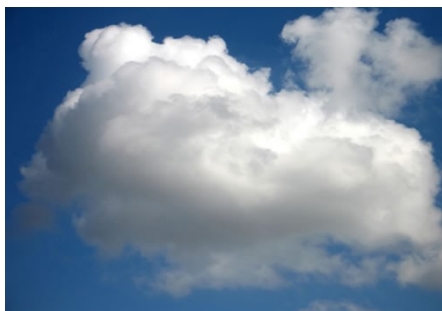
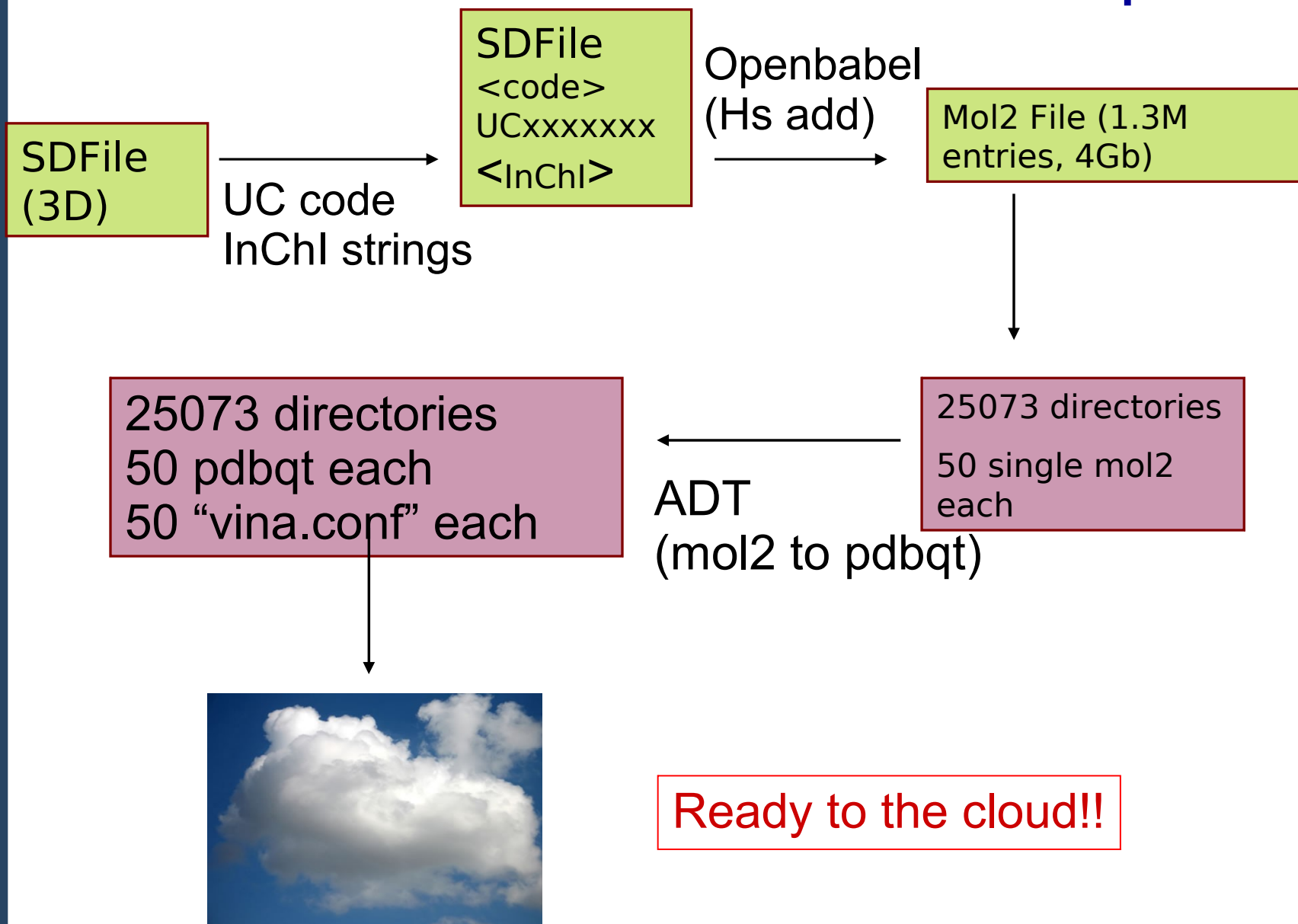


Chemical Database Processing

- **The Chemical database in U. of Cincinnati:**
 - Pipeline Pilot Generation of all possible xomers.
 - No filtering (look for pharmacological tools).
 - The database is provided as an SDFfile.



Data Base Preparation



Ready to the cloud!!

FEDERATED CLOUD FOR HPC

The Issue

- **Synergy chemical processing needs a HPC/HTC (High Productivity /High Throughput) cluster as big as possible to work properly.**
- **These resources are available at CESGA and FCSCCL centers (one center alone is not enough).**
- **Cloud Computing solves this issue joining distributed computing resources to work as a standalone HPC cluster.**
- **Applications requirements not suitable for static computing infrastructures:**
 - **OS requirements.**
 - **Software installation.**
 - **Jobs Management.**
- **Needs a “Custom” cluster solution.**

Hardware Resources

- **CESGA (Santiago de Compostela):**
 - 40 HP ProLiant SL2x170z G6. 2 Intel E5520 (Nehalem). 4 cores per processor. RAM 16 GB.
 - 1 HP ProLiant DL160 G6 2 Intel E5504 (Nehalem). 4 cores per processor. RAM 32 GB.
 - 1 HP ProLiant DL165 G6 2 AMD Opteron 2435. 6 cores per processor. RAM 32 GB.
 - 6 HP ProLiant DL180 G6. 2 Intel E5520 (Nehalem). 4 cores per processor. 16 TB de almacenamiento total.
- **FCSCCL (Leon):**
 - 32 Proliant BL2x220c. 2 Intel Xeon E5450. 4 cores per processor. RAM 16 GB.
 - 800 GB storage (NFS)

■ Features:

- VMs could be connected using a pre-defined “Virtual Network”.
- VMs could be started using a “golden copy” machine as reference.
- It's possible to define a different “context” for each executed VMs to modify the original “golden copy”.
- Could be defined a scheduling mechanism to select a specific physical host (based on round robin/ host load/ etc).
- It's possible to stop, start, migrate and save VMs.
- OpenNebula cluster could be used as HPC cluster (we manage Virtual Cluster VC instead of Virtual Machines).

Virtual Clusters

- **A Virtual Cluster (VC) could be used as a group of VMs:**
 - This VC includes a VM head node.
 - Several VMs are associated to VC head.
 - VC Virtual machines are interconnected using their own network.

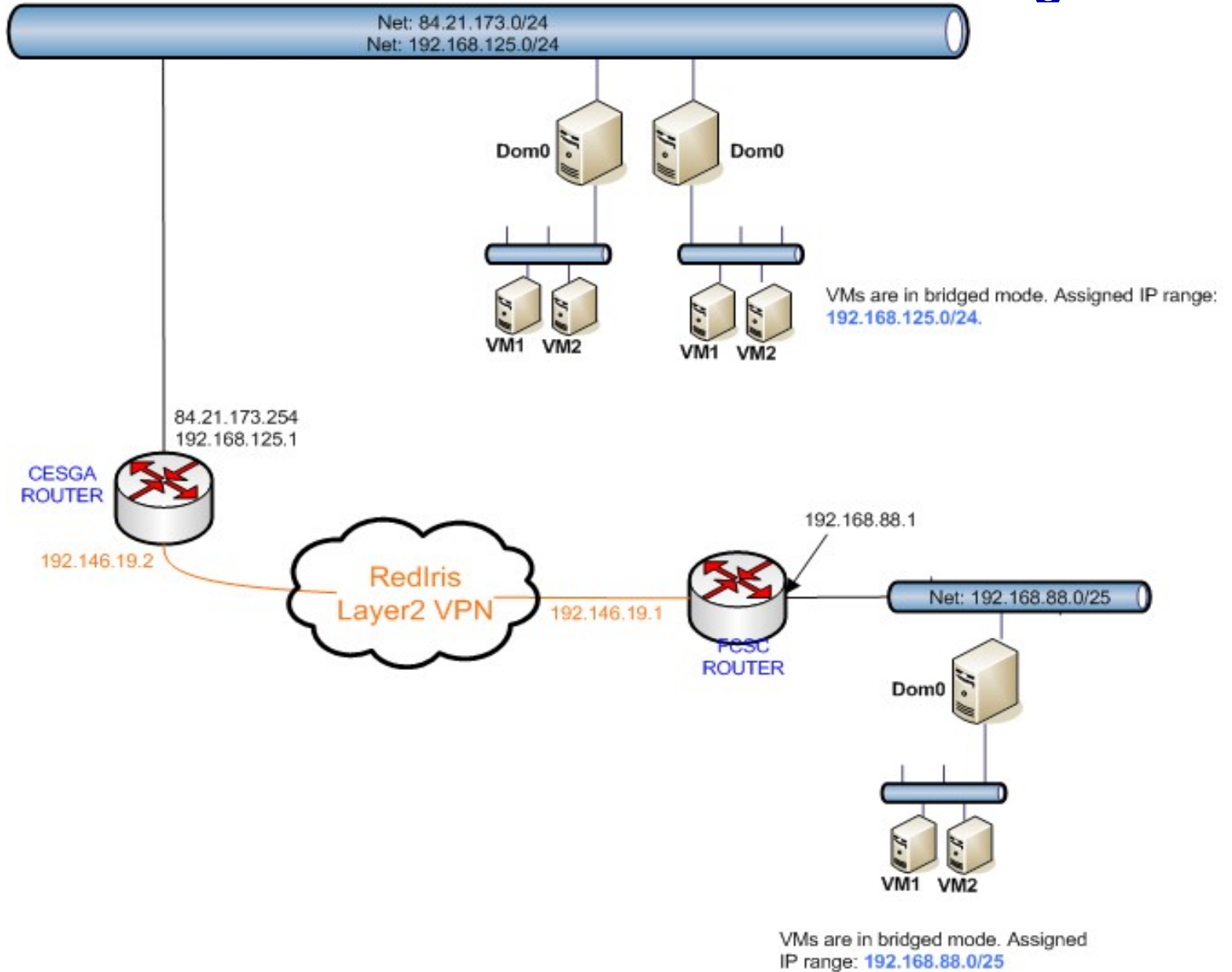
- **VCs are managed using different scripts:**
 - `make_cluster.sh`: To create a new VC. (Cluster name, network, nodes number, etc)
 - `kill_cluster.sh`: Delete VC. (Selects a cluster name to destroy).
 - `make_extra_node.sh`: To add cluster nodes.
 - `delete_n_nodes.sh`: Delete specific number of nodes.

- **VCs offers:**
 - Automated network configuration.
 - GE batch system is configured automatically with each VC creation.
 - Head node is not affected by VC nodes creation or destruction.

Network Configuration

- We need a “path” between resource centers (CESGA and FCFSL).
- OpenNebula server and the physical nodes must have a configured network routing.
- VC “head” must have public and private IPs.
- VC nodes are connected using a private network.

Network Configuration



OpenNebula Frontend

- User can connect to a web page to create or destroy VM.
- User also can use a private machines repository or store their own SO images.

The screenshot displays the OpenNebula Sunstone web interface. At the top, the navigation bar includes 'OpenNebula Sunstone', 'Documentation | Support | Community', and 'Welcome oneadmin | Sign Out'. Below the navigation bar, there are buttons for '+ New', 'Shutdown', 'Previous action', and 'Delete'. The main content area is divided into two sections. On the left, there is a table with columns for 'All', 'ID', and 'User'. The table contains several rows of data, including entries for 'oneadmin' users. On the right, there is a table with columns for 'Hostname' and 'Start Time', displaying a list of virtual machines with their respective hostnames and start times. In the center, a modal window titled 'Create a new Virtual Machine' is open. This window has tabs for 'Wizard KVM', 'Wizard XEN', 'VMWare', and 'Advanced mode'. The 'Wizard XEN' tab is selected. The wizard contains several sections: 'Capacity options' with input fields for 'Name', 'Memory', 'CPU', and 'VCPU'; 'Boot/OS options'; 'Add disks/images'; 'Setup Networks'; 'Add Graphics'; 'Add context'; 'Add placement options'; and 'Add Hypervisor raw options'. At the bottom of the wizard, there are 'Create' and 'Reset' buttons. The footer of the page contains the copyright information: 'Copyright 2002-2011 © OpenNebula Project Leads (OpenNebula.org). All Rights Reserved. OpenNebula 2.1.80'.

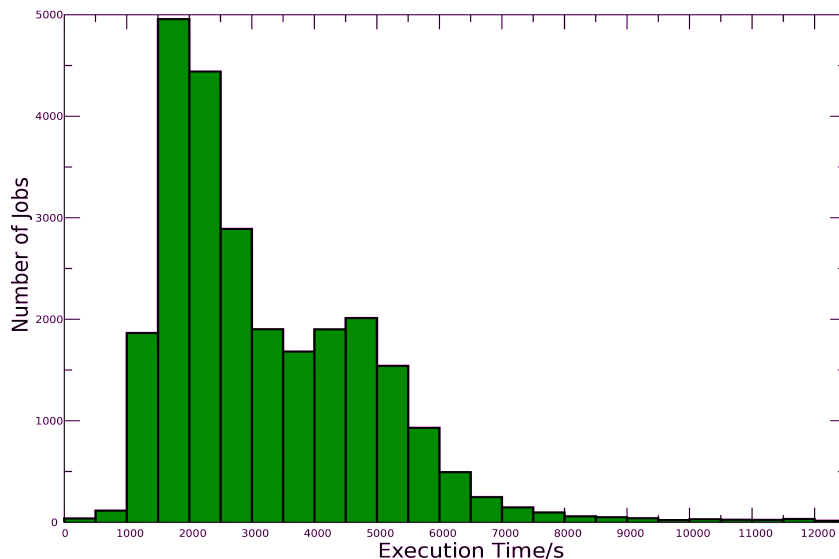
Experiment Results

- Job execution was started on August 15.
- Finished on September 31.

	Used Cores	Total Execution Time/s	Total Jobs	Average Job execution time/s	Efficiency (%)
VINA	322	1214530	25690	3412	22.4
VSW	64	331016	191	96390	86.9

VSW already has a efficient job manager

Vina: 131 jobs exceed 12500 s.
Some jobs reach near 700000 s.



Vina supports SMP parallelization
+
Efficient job grouped algorithm is needed



Efficient vina job manager to be developed

CONCLUSIONS

Conclusions

- **Cloud Computing techniques allow to test VCs in a short period of time.**
- **Deploy VCs is faster than a physical cluster installation.**
- **Ad-hoc clustering for different users need (SO, Software, etc).**
- **And Its maintenance consumes less manpower and time.**
- **Users can administrate their own virtual machines using VCs.**
- **VC “head” must have public and private IPs.**
- **It's possible to create geographical distributed VCs.**

**THANK YOU FOR YOUR
ATTENTION!**

¿Questions?