

THE DECRYPTHON PROJECT : HELPING CURE MUSCULAR DISTROPHY THROUGH GRID AND VOLUNTEER COMPUTING

David Loureiro – CEO@SysFera

Open World Forum – FOSS For Humanity

11 octobre 2012



Agenda

- SysFera
- The Decryphon project
- The platform architecture – volunteer computing
- The platform architecture – grid computing
- HCMD : an application deployed on both platforms
- The future

SysFera

- We provide FOSS to help Large IT infrastructure owners and ISVs adopt a SaaS delivery model
- Based on research software called DIET
- 14 people located close to the INSA campus in Lyon (the weirdos doing pixel art on the windows? That's us.)
- Références:



The Décryphon project

- 2001 : Launched by AFM/Téléthon in conjunction with IBM and CNRS
- Goal : Helping Cure Muscular Dystrophy through in-silico simulations
- Two actions:
 - One long-run project with a lot of individual tasks that could run on a large number of individual PCs through volunteer computing
 - Several project needing HPC resources to run managed by researchers on a dedicated grid environment



The volunteer computing platform

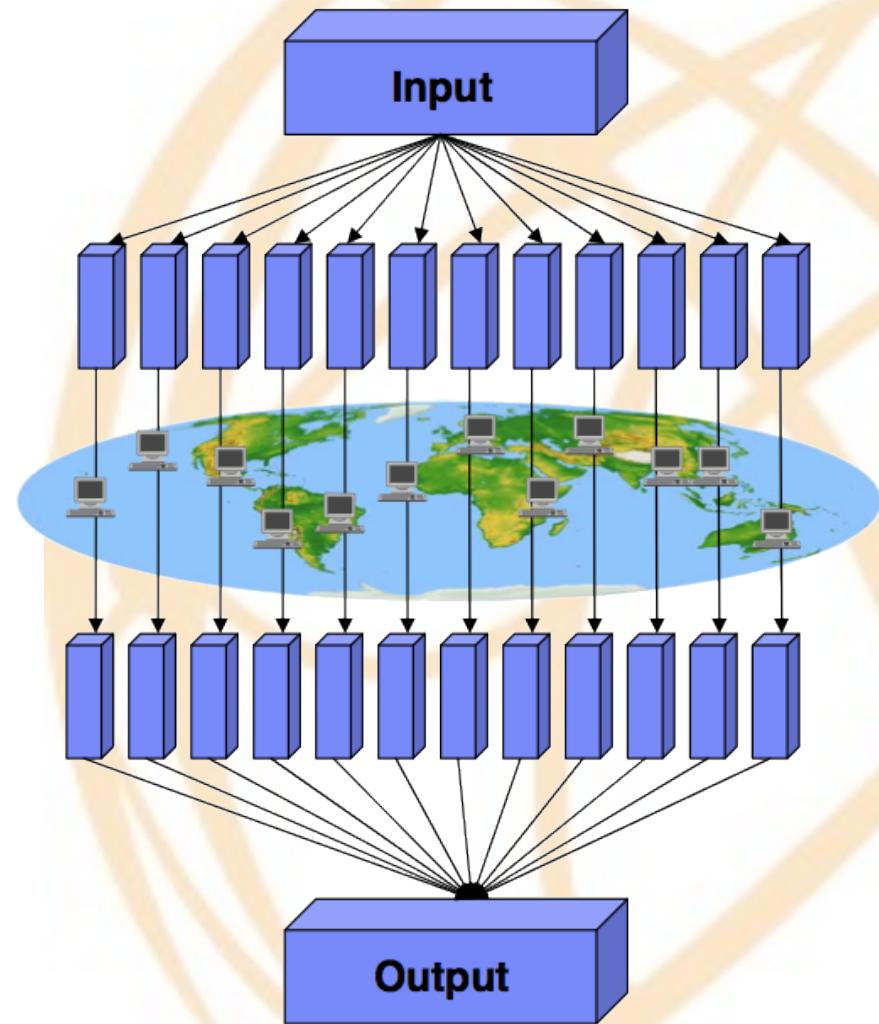
- Framework : World Community Grid (launched in 2004 by IBM)
- Several projects :
 - Human Protein Folding
 - FightAIDS@home
 - Help Defeat Cancer
 - Genome Comparison



Crédit : IBM

The volunteer computing platform

- Typical project operation :
 - Humanitarian project selected which requires massive amounts of computing to solve
 - Large computing problem is split into millions of smaller independent runs
 - Member PCs ask for a piece of work to run and then send back the results that will be assembled to get the final results
 - Direct results made available in the public domain
- **PERFECT MATCH !**



Crédit : IBM

The volunteer computing platform

- [Project] Preparation steps to enter World Community Grid
 - RPF Submitted, evaluated and approved
 - Code is prepared by researchers following guidelines
 - And then validated by IBM before being proposed to members
- [USER] Classical workflow
 1. Member joins,
 2. Download & install agent software from web site
 3. Register
 4. Agent asks for work
 5. Device processes work at lowest CPU priority
 6. Device contacts the servers
 7. Go to 4

The volunteer computing platform



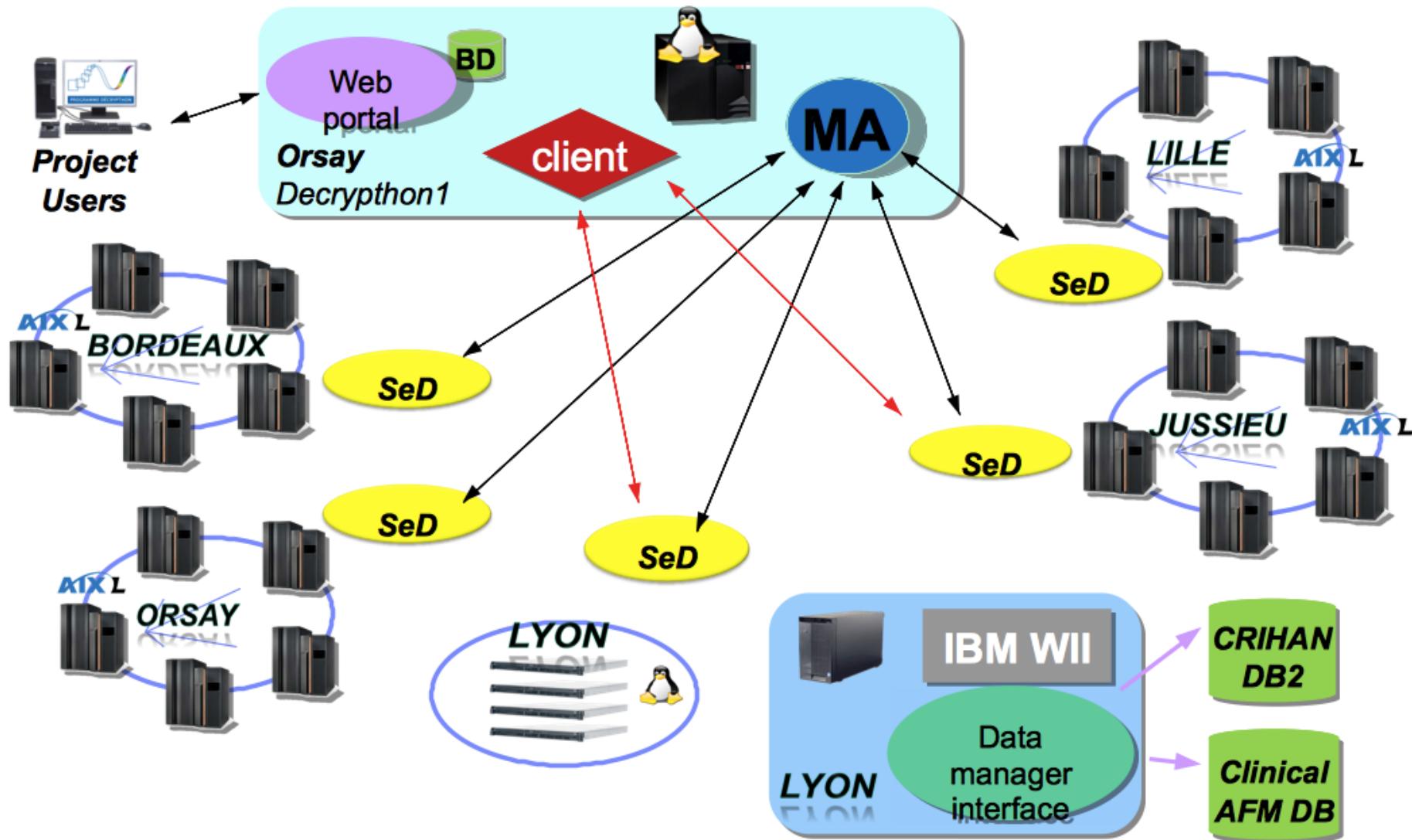
LGPL Licence

- Based on Boinc -> Developed @ Berkeley University
- Boinc is **secured** from both user and project point of view
- Projects are **independent**; There is no central directory or approval process.
- Boinc is used for several projects : SETI@Home, Rosetta@Home, Mersenne@Home
- **BOINC Stats: +2,4M users in +250 countries**
- **WCG Stats : ~400k users on ~1,6 computers in +220 countries**

The grid computing platform

- Framework : strong collaboration between CNRS/AFM/IBM and ENS Lyon
- The DIET Middleware will be used to provide a grid used by researchers to perform computations based on HPC applications, huge amount of data and more specific research topics

The grid computing platform



The grid computing platform

- Application-centric interface were available to execute simulation
- Zero knowledge of networking, data base, processing, etc needed

Elapsed time, Number of results and Result file size repartition
From 2008-02-29 00:00:00 to 2008-02-29 16:14:22

Pie chart of the elapsed time (Hours) : device repartition

Device	Elapsed Time (Hours)
Breggat	6,355
gdsm10	0,407
gdsm11	0,313
gdsm12	0,305
gdsm13	2,378
gdsm14	2,543
gdsm15	18,381
gdsm16	11,384
gdsm17	0,888
gdsm18	0,479
gdsm19	0,894
gdsm20	0,377
gdsm21	2,793
gdsm22	6,559
gdsm23	12,253
gdsm24	11,822
gdsm25	12,343

annotation
1 job, submitted with web interface. Date : 15/1/2008

ign v1.0

Blast	
expect	5000
max alignments	5000
search	5000
Databank	ptb
f value	9
F value	<input checked="" type="checkbox"/>
Filter	
threshold	0.001
max seq kept	500
length	5000
Filtering method	none
last seq	<input checked="" type="checkbox"/>
remove fragments	<input checked="" type="checkbox"/>
Clustal	
ufs	<input checked="" type="checkbox"/>
propagate	<input checked="" type="checkbox"/>
ject name	NULL
debug info	<input checked="" type="checkbox"/>

List of the permissions for all users

User name	Administrator	Users management	Devices management	Statistics viewer	logging	ms2ph	spikeomatic	Default
root	<input checked="" type="checkbox"/>							
nbard	<input checked="" type="checkbox"/>							
raphael	<input checked="" type="checkbox"/>							
Spikesomatic user	<input checked="" type="checkbox"/>							
MS2PH user	<input checked="" type="checkbox"/>							
statistic_viewer	<input checked="" type="checkbox"/>							
DOCKING user	<input checked="" type="checkbox"/>							

DIET Webboard for the Decryphon project

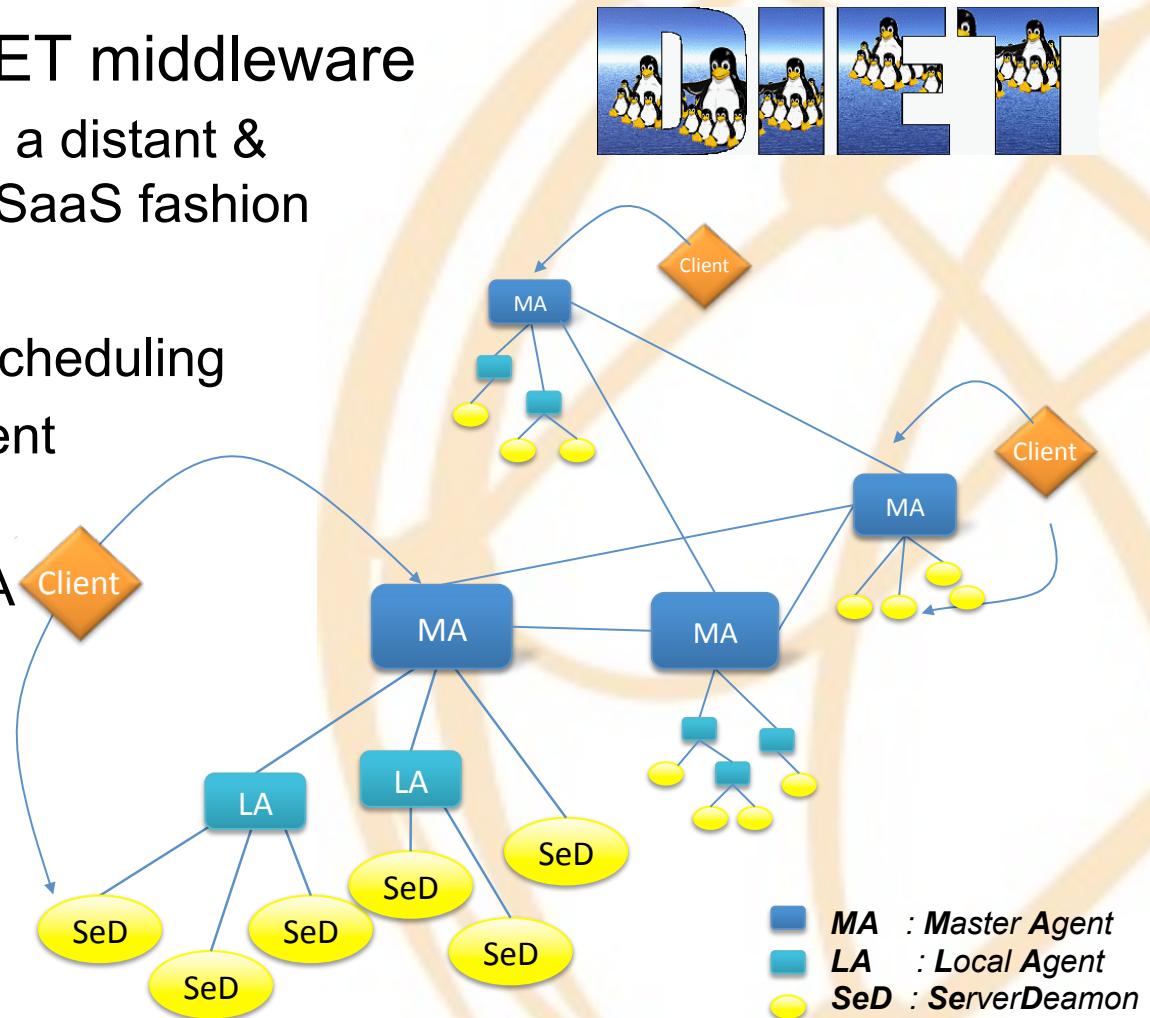
DIET Daemon manager	Application List	View generated GoDIET xml
DIET properties	Data List	Installed applications on devices
My account	Device List	Grid Statistics
View users permissions	Device group List	
Manage temporary files	Job List	
View DIET client log files	Request List	View running jobs
Write a news	Result List	View running workunits
DIET MS2PH Demonstration	Storage List	
DIET Docking Demonstration	User List	
DIET SpikeOMATIC Demonstration	Workunit List	
	News List	News from decryphon
		Back to the identification menu

download all available results of this job

Home page

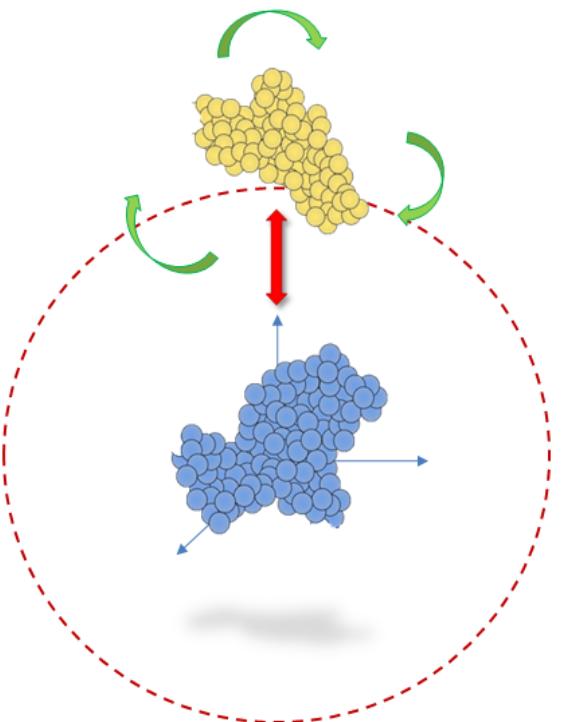
The grid computing platform

- The power of the DIET middleware
 - Access application on a distant & optimized server in a SaaS fashion
 - Fault-tolerant
 - Provided distributed scheduling
 - Workflows management
 - Data management
 - Open source: CeCill-A



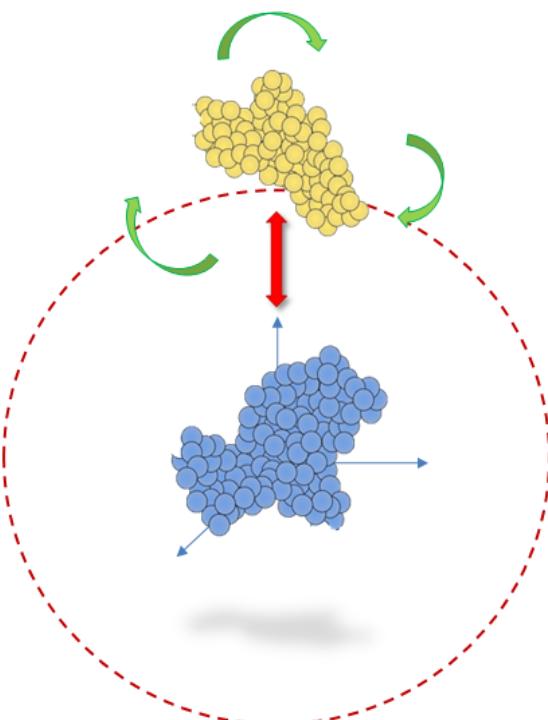
HCMD : app deployed on both platforms

- The Docking project:
 - Also known as « Help cure muscular dystrophy » the aim is a large scale investigation of protein-protein, DNA-protein and protein-ligand interactions in the search of new therapeutic targets. It is done by developing computing tools able to locate on the surface of proteins, interaction sites with DNA, ligands, and other proteins.
 - Coordinated by Alessandra Carbone from the "université Pierre et Marie Curie" , Paris (Inserm U511 - Immunologie cellulaire et moléculaire des infections parasitaires - Génomique analytique)



HCMD : app deployed on both platforms

- Launched on May 12, 2009
- Status : 98,.. % done
- Results:
 - 113 million results processed
 - 53,000 years of computing power
 - Research completed in 2,5 years



The future of the Decrypthon project

- IBM, the CNRS and SysFera (supporting and co-developing the DIET middleware) are working on the next Decrypthon generation
- Goal : provide a Cloud infrastructure to continue the research on Muscular Distrophy
- The CFPs have been sent
- Machines will be installed before the end of the year in computing centers
- There will never be enough initiatives to help researchers curing muscular dystrophies that Big Pharma don't find « profitable »

THANKS !

@DavidLoureiroFr <david.loureiro@sysfera.com>

@SysFera <contact@sysfera.com>

