CARRIOCAS
Collaborative High Performance Scientific Visualization

Christophe MOUTON, EDF R&D
Jean-Philippe NOMINE CEA/DIF/DSSI
With contributions of : ECP, OXALYA

AGENDA

- Introduction
- Part 1 : Goals of the « Collaborative High Performance Scientific Visualization » demonstrator of the CARRIOCAS project
- Part 2 : The first results of the CARRIOCAS teams
- Part 3 : Future work

Introduction

CARRIOCAS ?

CARRIOCAS

- **« Distributed Computation Over Ultra High Speed Optical Internet Network »**
- **In french : « CAlcul Réparti sur Réseau Internet Optique à CApacité Surmultipliée »**

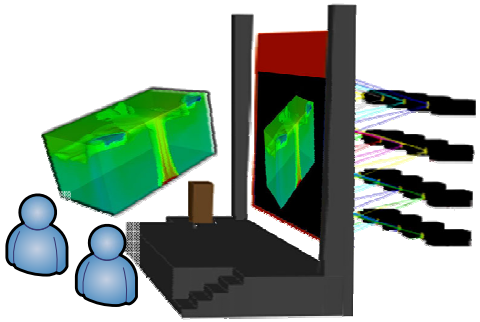
- A 3-year project in the frame of the French SYSTEM@TIC Competivity Cluster: Oct. 2006 – Sept. 2009

SYSTEM@TIC PARIS-REGION Pôle de compétitivité

CARRIOCAS : Ambition

- **40 Gbits to model and simulate en real time**

  - **Adaptation of optic techniques to reach this ultra high bit rate**
  - **Integration and validation on an experimental network at the top level bit rate of 40Gb/s**
  - **R&D for applications development:**
    - **Distributed storage of massive data on remote servers**
    - **Remote Collaborative High Performance visualisation**



Partnership

CARRIOCAS

Project leader
WP (40Gb/s transmission) and
WP 2 (protocols and network architecture)

Academics

91

92

78

75

Alcatel·Lucent

III-V lab
ALCATEL-THALES

Draka Comteq

hp invent

M.E.D.I.T

INRIA

Supélec

UNIVERSITÉ PARIS-SUD 11

WP4 leader
(Distributed applications)

CNRS

LPN

176 people.year
Oct 2006-Sept 2009

Atos Origin

EDF

LIST

CENTRALE PARIS

RENAULT

UNIVERSITÉ DE VERSAILLES SAINT-QUENTIN-EN-YVELINES

PRISM

kylia

Oxalya

N2Nsoft

get

WP 3 leader
(experimental network)

Partenaires financeurs:

MINISTÈRE DE L'ÉCONOMIE DES FINANCES ET DE L'INDUSTRIE

MAIRIE DE PARIS

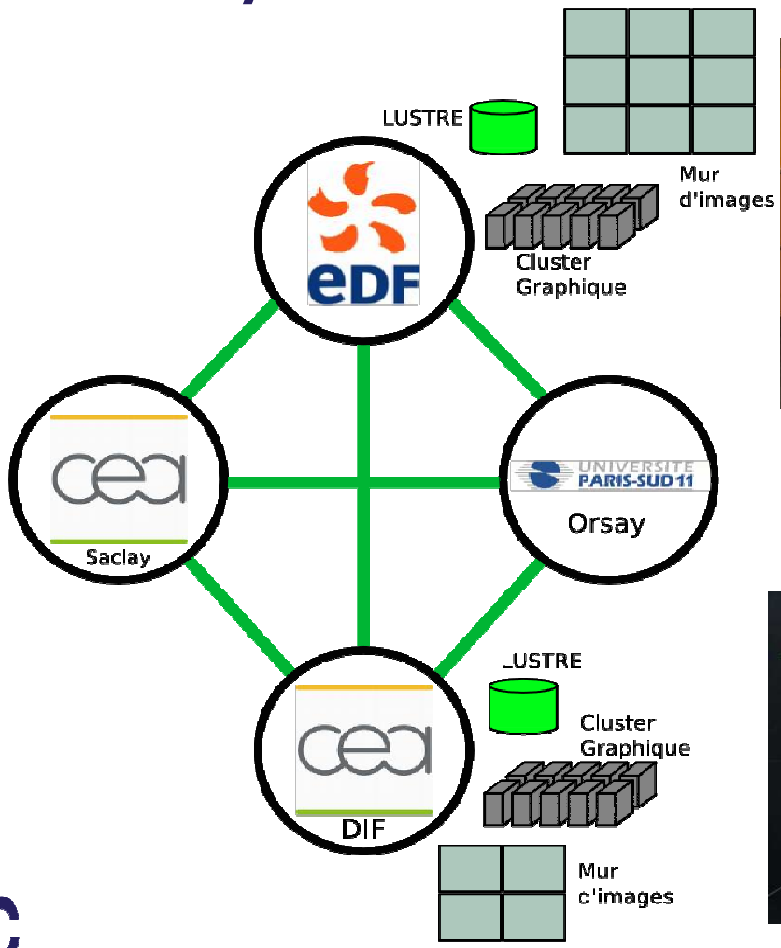
Essonne LE CONSEIL GÉNÉRAL

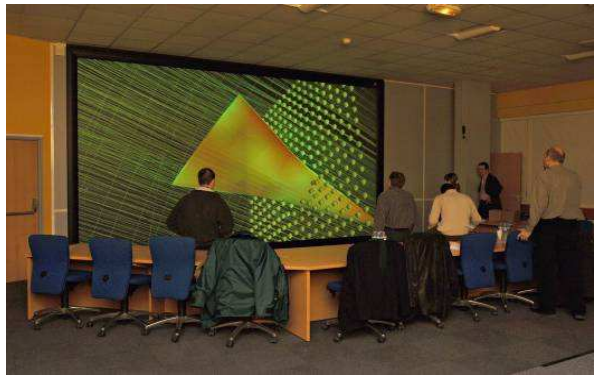
92 CONSEIL GÉNÉRAL DES HAUTS DE SEINE

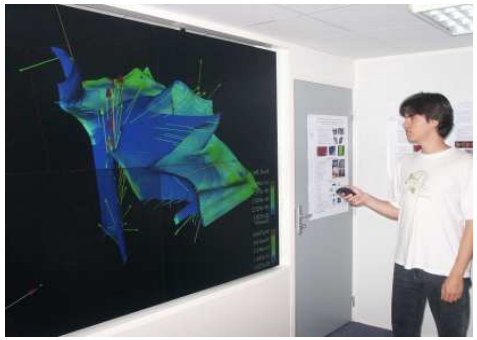
Distributed Computation Over Ultra High Speed Optical Internet Network

CARRIOCAS in 3 lines

- A Distributed Massive Filesystem (LUSTRE)
- Remote High Performance Visualisation
- **Over a 40 Gb/s Network**







Part 1 : Goals of the « Collaborative High Performance Scientific Visualization » demonstrator of the CARRIOCAS project

Scientific and Industrial context

Focus on the daily use of HPC

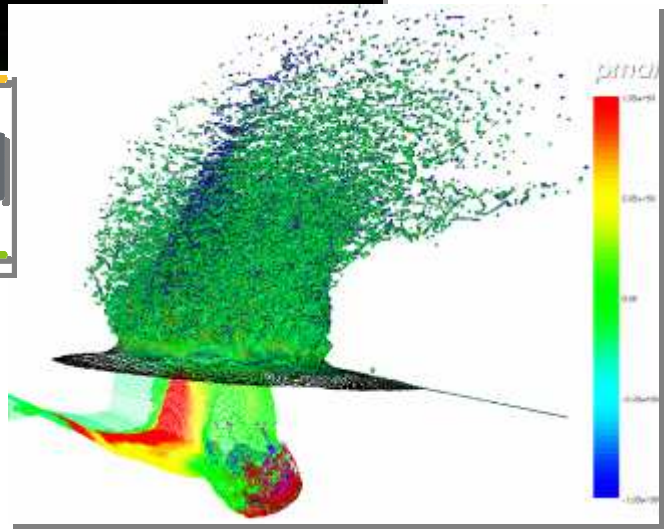
inside a Power Utility : EDF

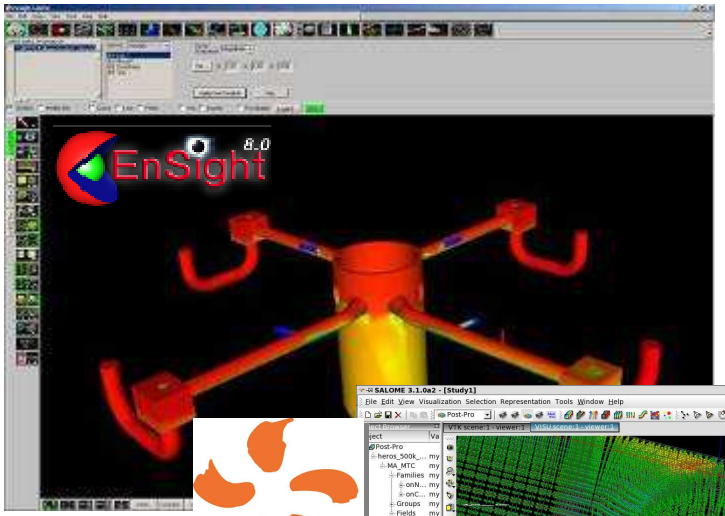
Scientific and industrial Context 1/2

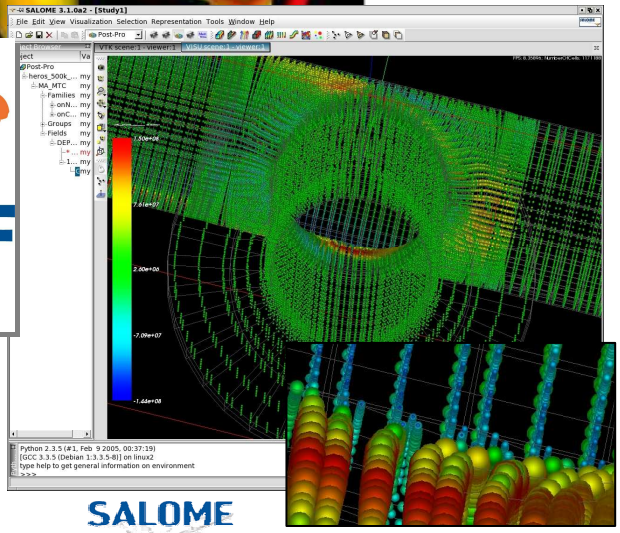
- High Performance Computing :
  - Examples of recent HPC ressources by the CARRIOCAS Partners
    - CCRT : Bull, Platine 47 Tflops, 26th top500
    - EDF R&D : IBM Bluegene/L : 22,9 Tflops, 53d top500

Scientific and industrial Context 2/2

- HPC : a compulsory tool for deep challenges :
  - Defense, Energy and Research : **CEA**
  - Energy of today and tomorrow : **EDF**





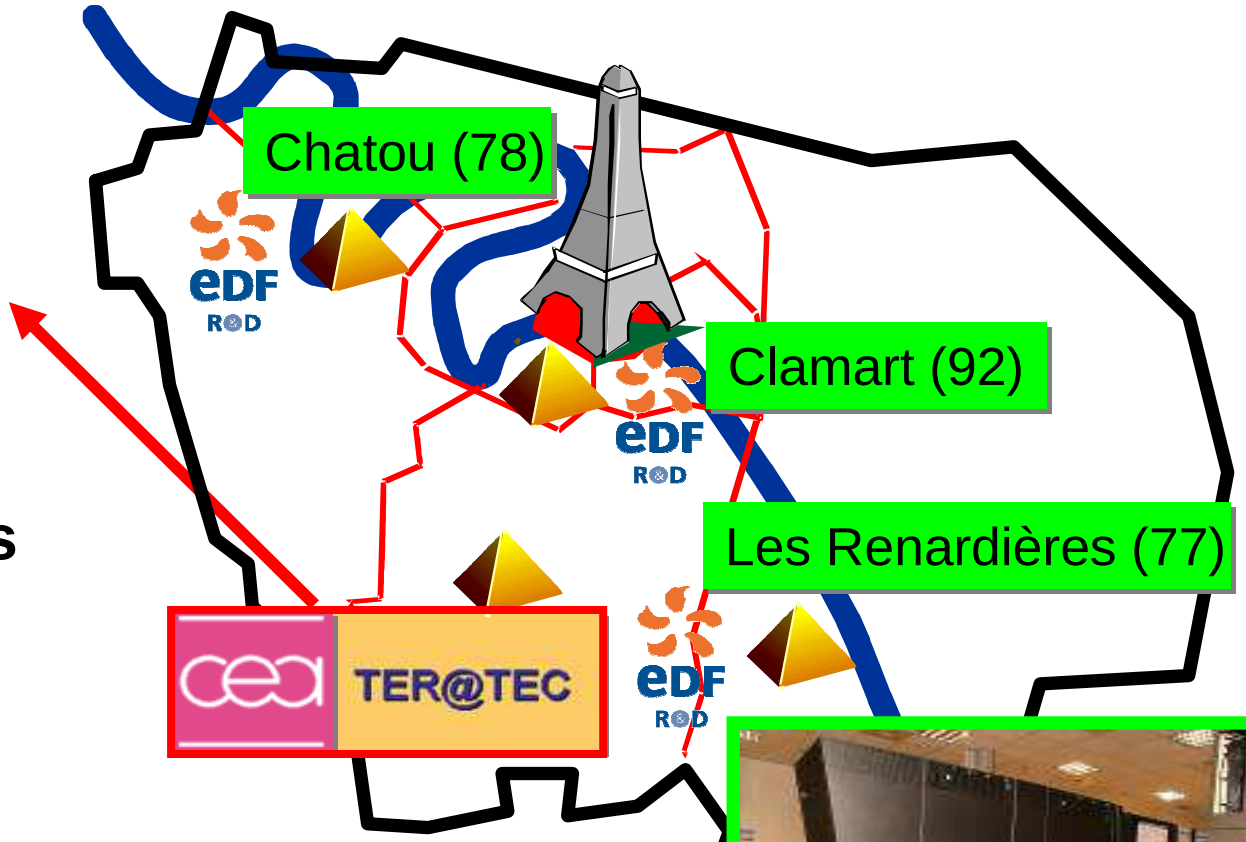


REX of HPC use in a Power Utility : EDF

EDF R&D : 600 researchers using HPC ressources

09/07 CCRT-B 43 TFlops
EDF Use = ¼ CPU Time

EDF HPC Power Use
2003 : 0,4 TeraFlops
2004 : 1,5 TeraFlops
2005 : 2,5 TeraFlops
2006 : 17,5 TeraFlops
2007 : 39,5 TeraFlops

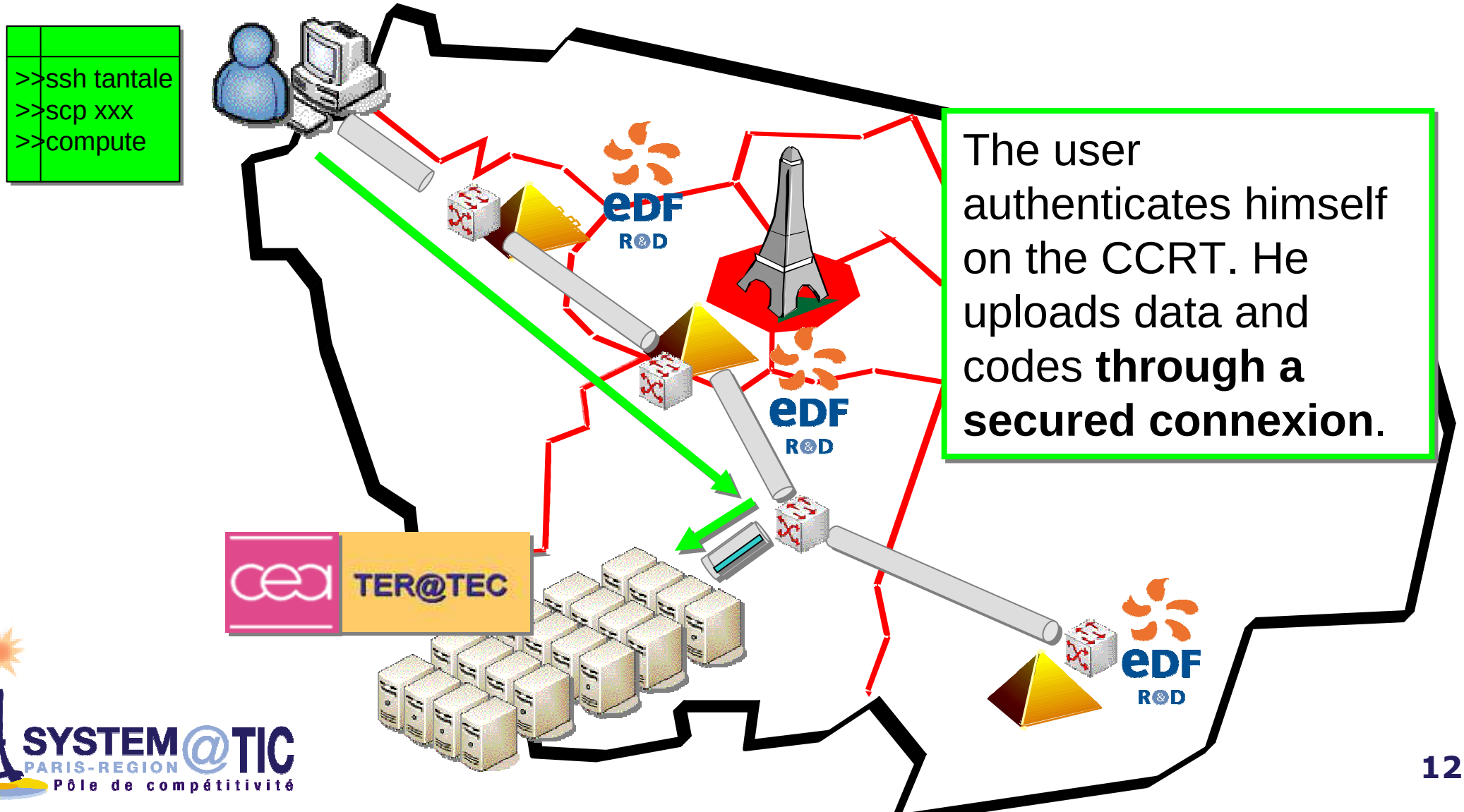


IBM BlueGene/L
1st european industrial supercomputer

SYSTEM@TIC PARIS-REGION Pôle de compétitivité

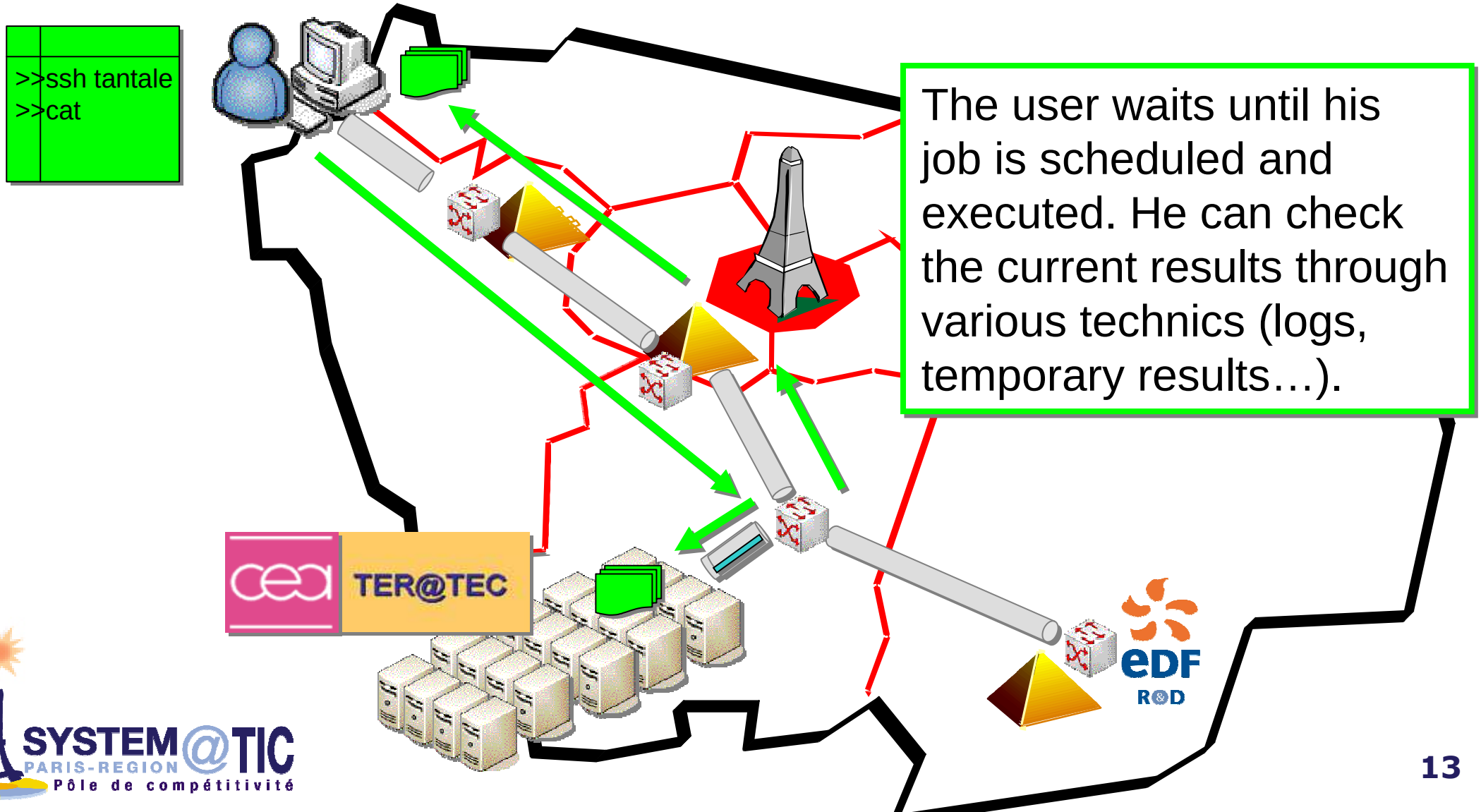
Today EDF HPC Use Case

- **An example from the CFD world : everyday use of the CCRT HPC ressources**



Today HPC Use Case

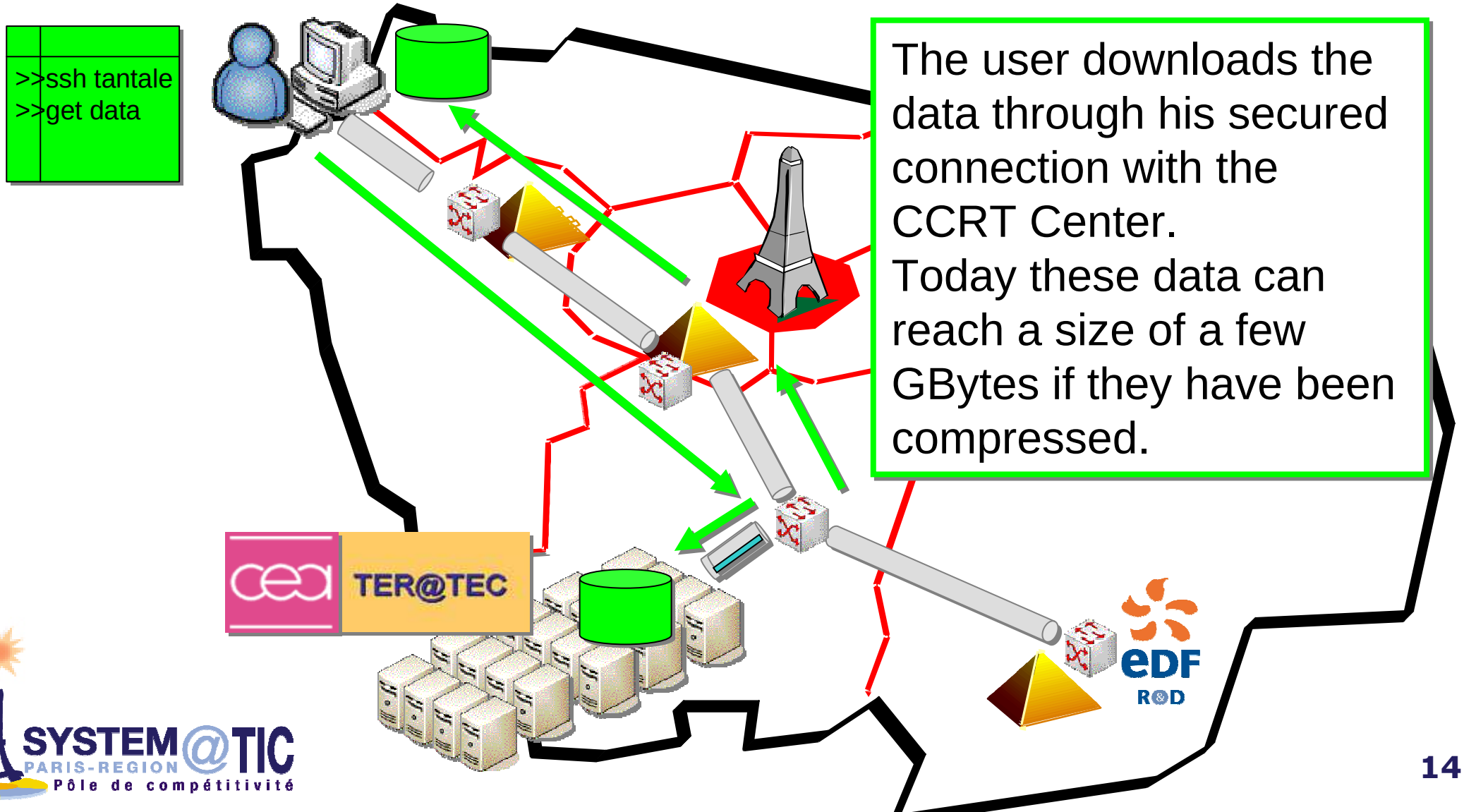
- Computation monitoring



The user waits until his job is scheduled and executed. He can check the current results through various technics (logs, temporary results…).

Today HPC Use Case

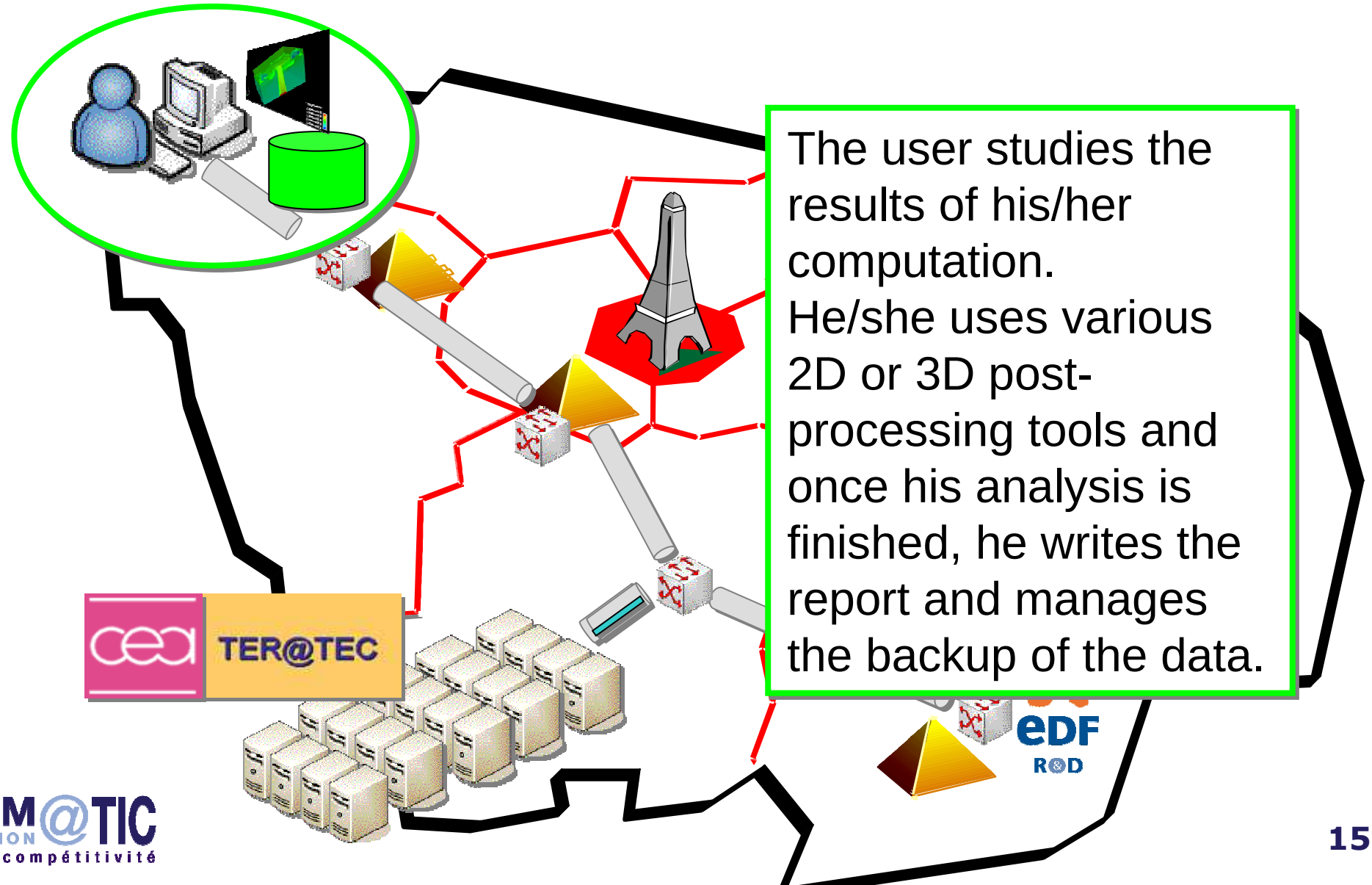
- When the computation is finished…



The user downloads the data through his secured connection with the CCRT Center.
Today these data can reach a size of a few GBytes if they have been compressed.

Today HPC Use Case

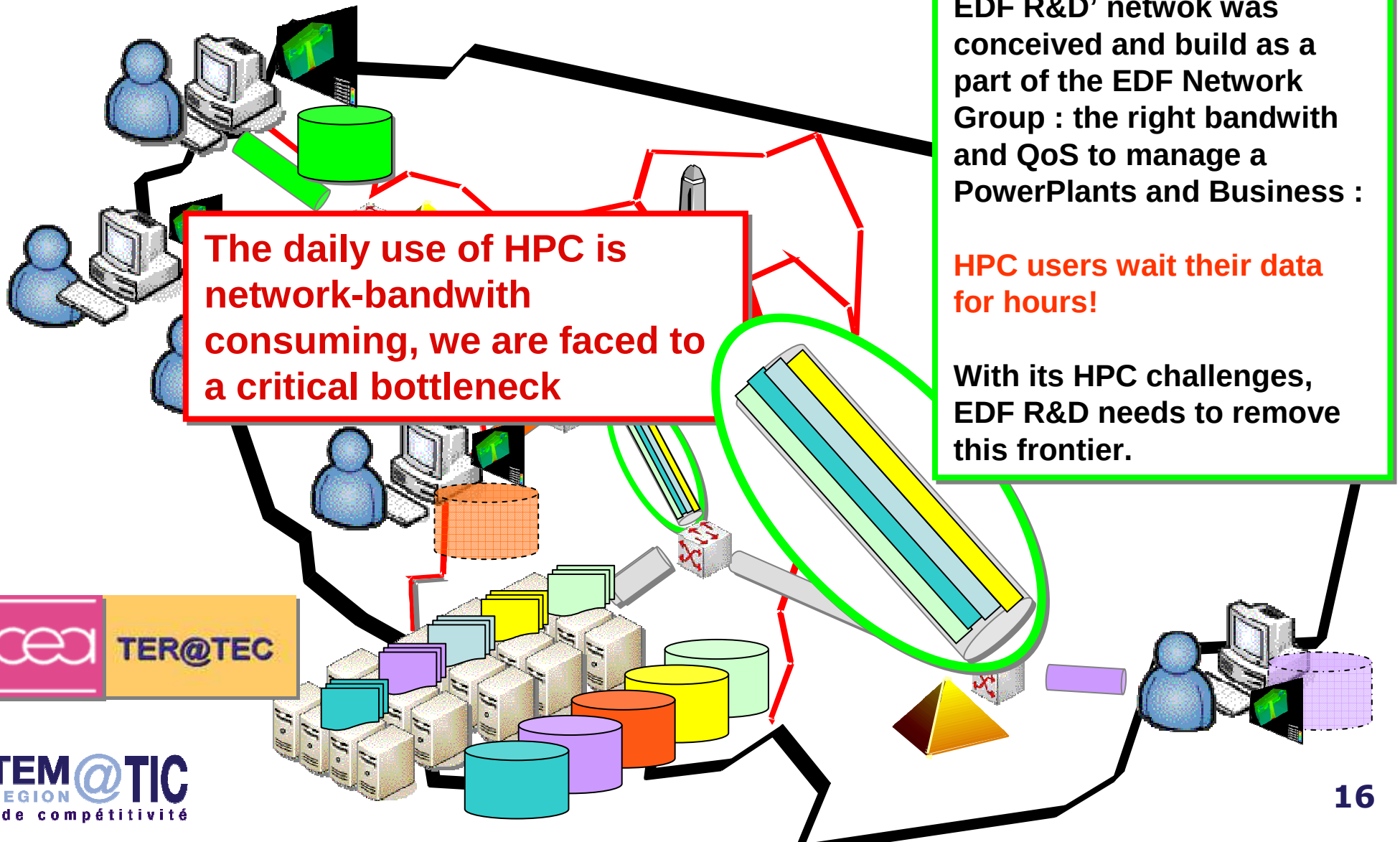
- Post-Processing on the user workstation



The user studies the results of his/her computation.
He/she uses various 2D or 3D post-processing tools and once his analysis is finished, he writes the report and manages the backup of the data.

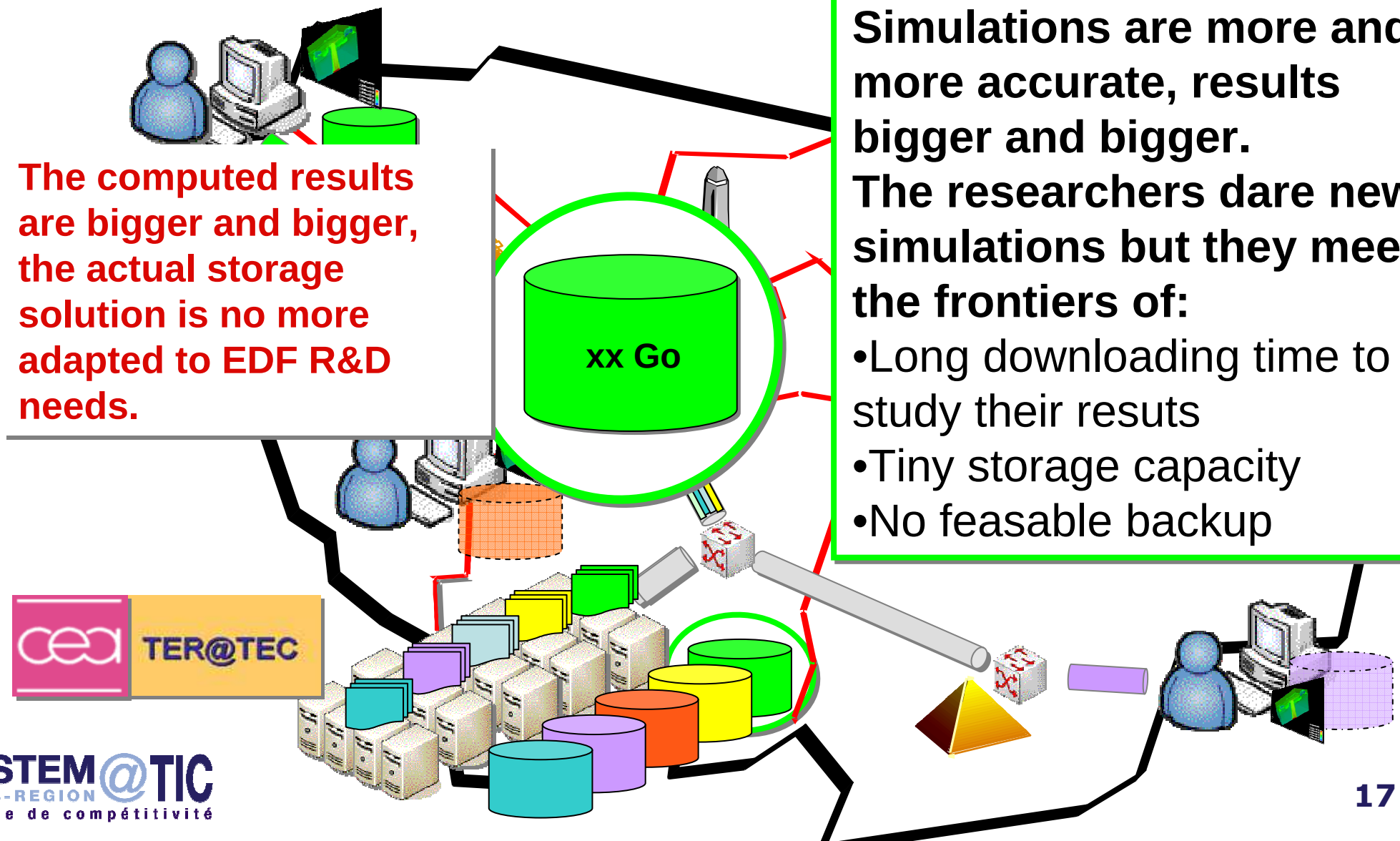
But we met new frontiers (1)

- This is one example but daily activity at EDF R&D is around 600 such scenarii



But we met new frontiers (2)

- Hom to manage these data?



The computed results are bigger and bigger, the actual storage solution is no more adapted to EDF R&D needs.

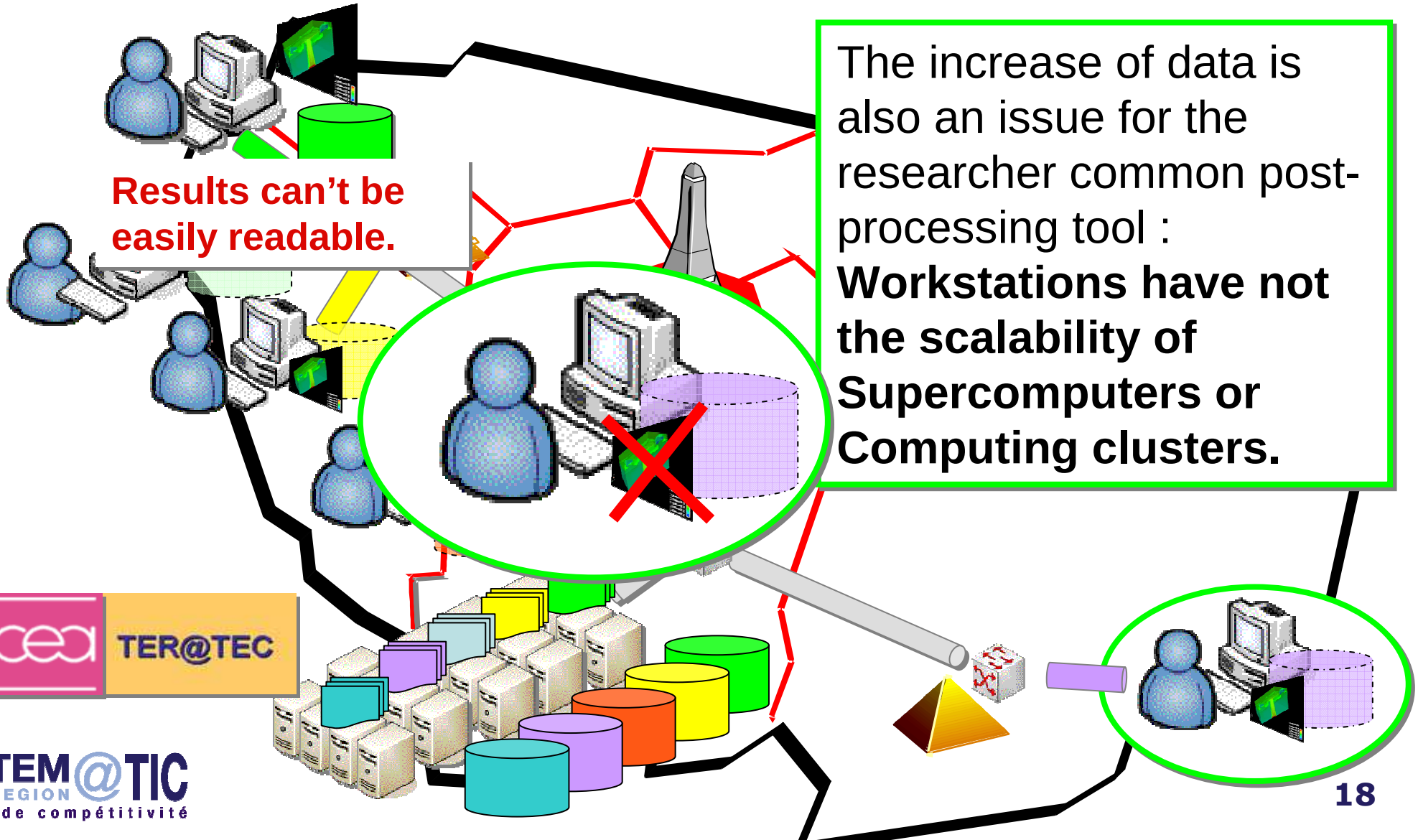
Simulations are more and more accurate, results bigger and bigger.
The researchers dare new simulations but they meet the frontiers of:
- Long downloading time to study their resuts
- Tiny storage capacity
- No feasable backup

TER@TEC

SYSTEM@TIC PARIS-REGION Pôle de compétitivité

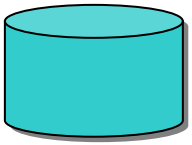
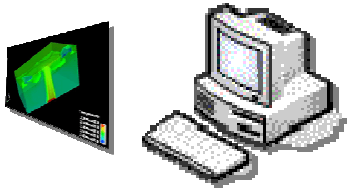
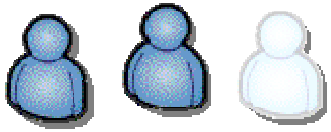
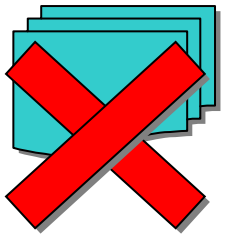
But we met new frontiers (2)

- How to analyse these data?



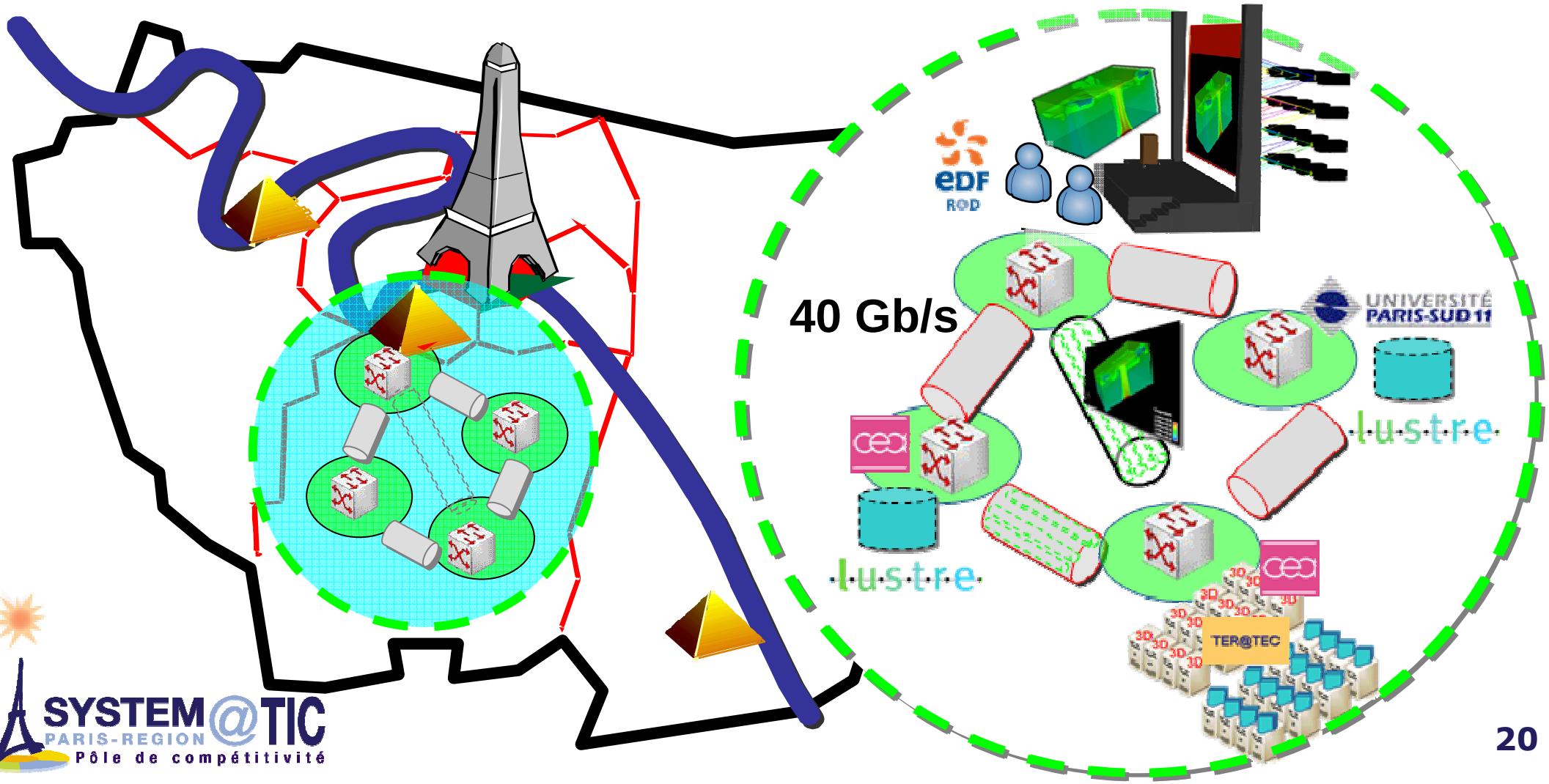
Technological Bottlenecks

- To manage dozens of GBytes of simulations data for one user among hundreds implies to design and deploy a global solution to solve the problems of :
  - **Networks (QoS, Bandwith..)**
  - **Storage and Backup**
  - **Analysis software and hardware ressources**
  - **Remote Collaborative expertise?**
  - Realtime Simulation Monitoring?

A provocative proposal : the CARRIOCAS technical architecture

- **CARRIOCAS : 40 Gb/s for**
  - A Distributed Massive Filesystem (LUSTRE)
  - Remote High Performance ScientificVisualisation

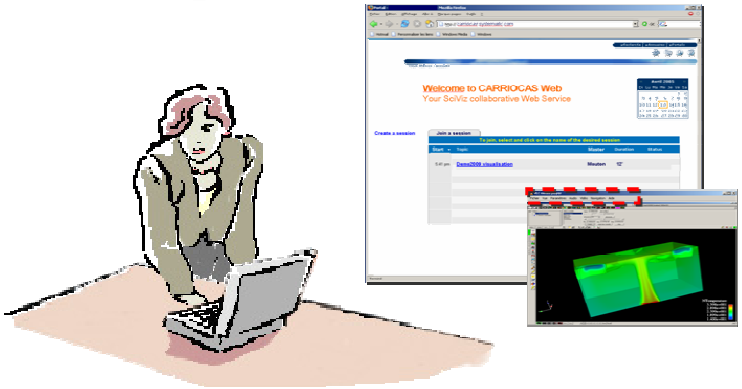


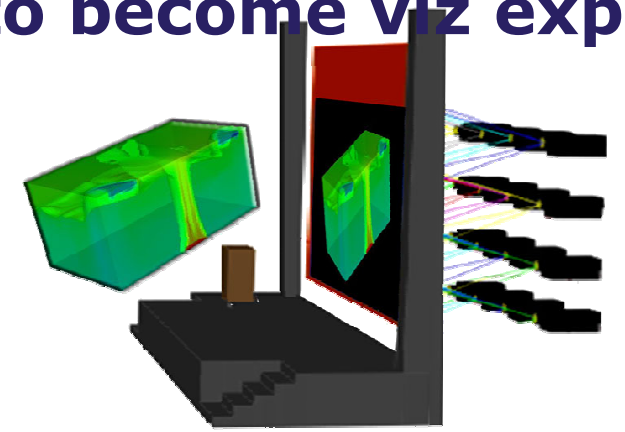
Target

- **Which kind of users? All of them !**
  - Researchers : physicists, mathematicians, numerical analysis specialists
  - Experts and Engineers of EDF engineering Units for :
    - CAD, safety studies

  Everybody who use daily
    - HPC tools

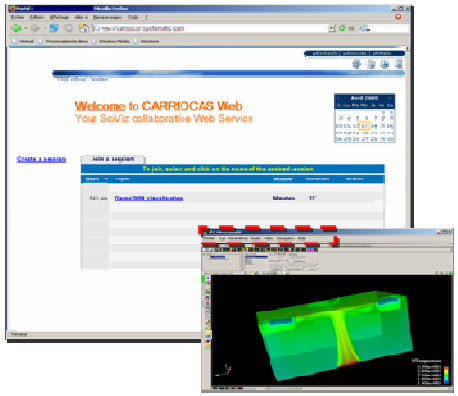
- **But new post-processing tools have to become User Friendly, user must not have to become viz experts**

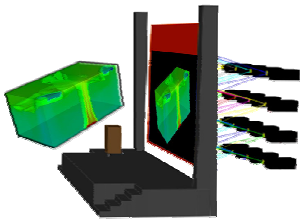




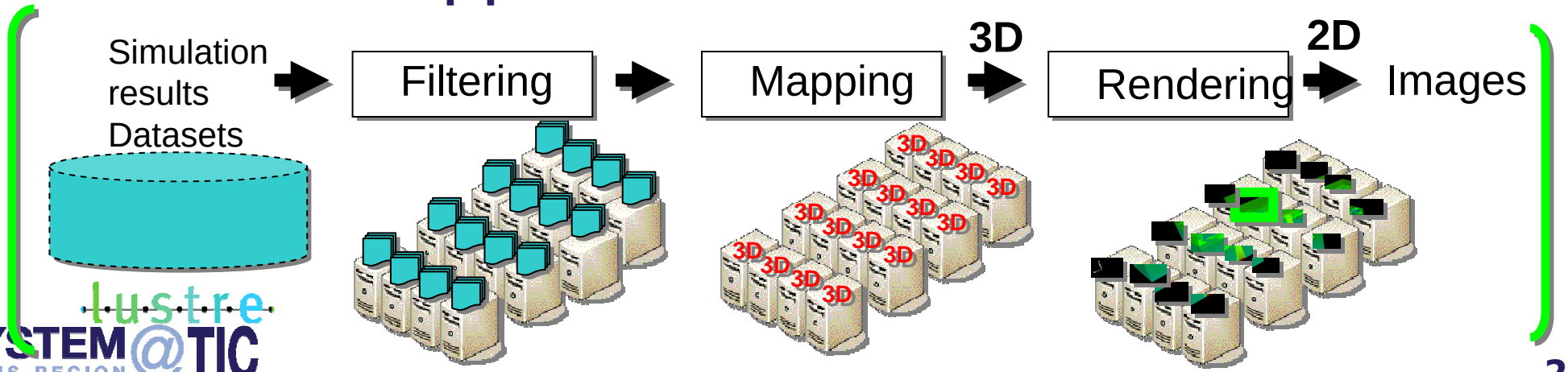
CARRIOCAS objectives:

- A new use of High Performance Scientific Visualisation:
  - **The Convergence of two worlds : IT and HPC**
    - **Thanks to the easiest way to access to scientific visualisation services : VISUPORTAL**





    - **Automatic configuration of HPC ressources for the scientific visualisation pipeline**



Part 2 : The first results of the CARRIOCAS teams : T0 + 12 month

The first VISUPORTAL prototype

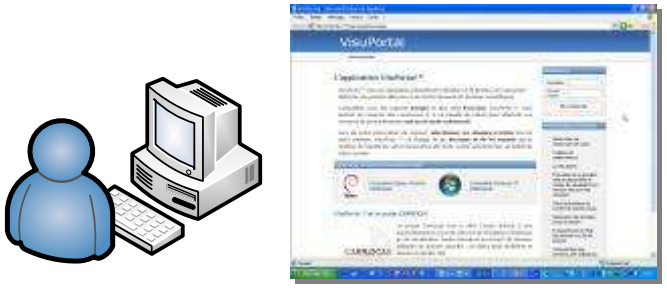
First experimentions

First technical results for :

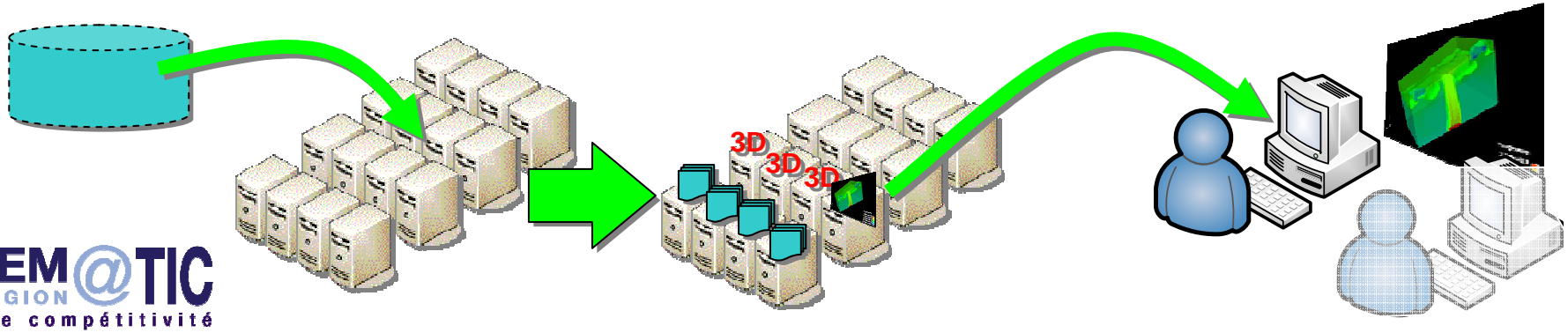
Lightweight streaming client (VLC), Scalability of Post-Processing techniques for large 3D scenes and High Resolution Displays

The VISUPORTAL prototype

- A **VISU**alisation **PORTAL** by OXALYA
  - For booking and using Remote High Performance Visualisation ressources
  - With a first integration of a remote visualisation tool : HP Remote Graphics (HP RGS)
  - Conceived and tuned for the post-processing software EnSight with its various versions : standard, Gold, DR (Distributed Rendering)
  - Equiped with monitoring ressources : RAM, CPU, Network uses (HURRICANE © OXALYA)

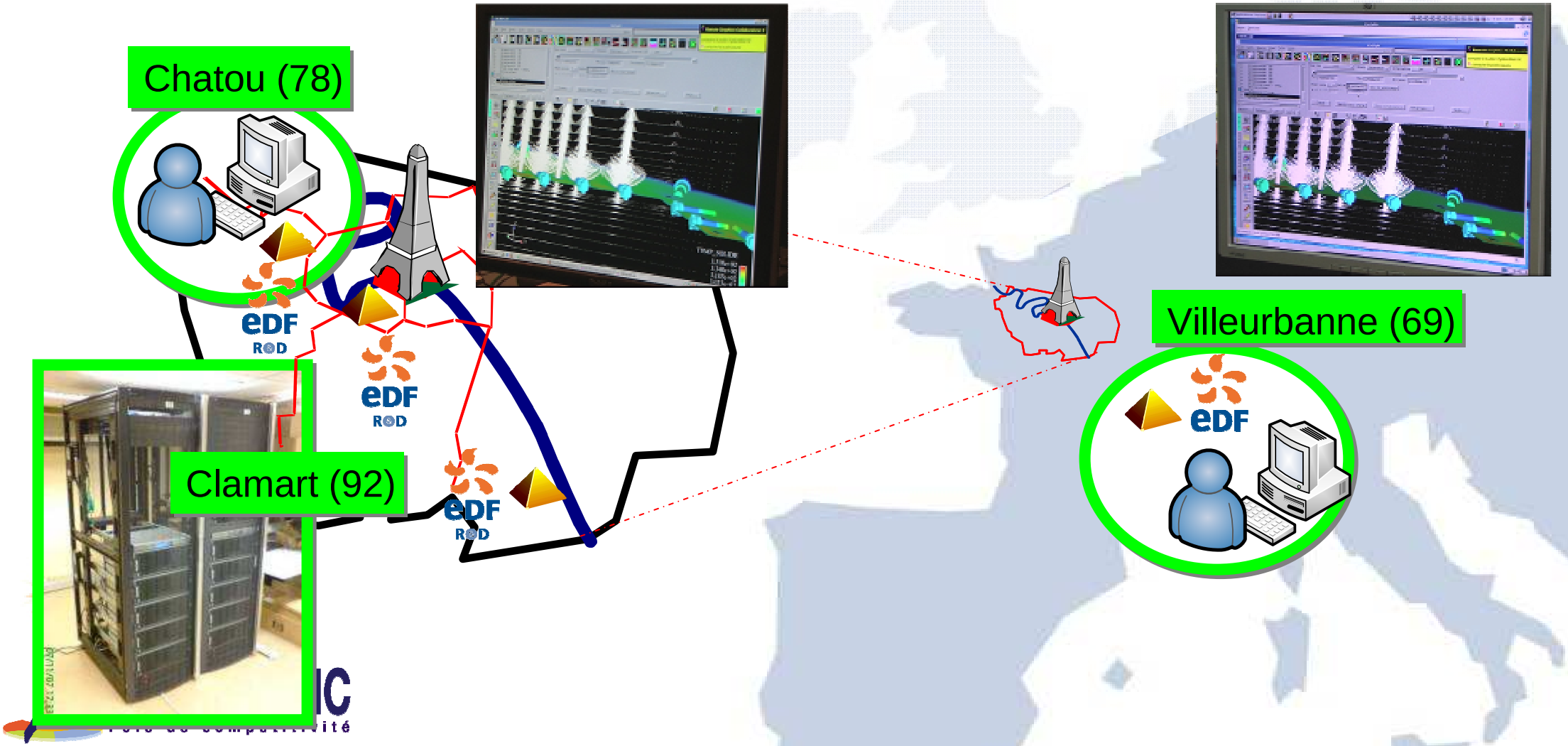


- The User
  - Selects one of his datasets
    - **VISUPORTAL checks the data**
    - **Allocate the corresponding right ressources**
  - Invite (or not) his colleagues
  - Then joins the session and his job begins.



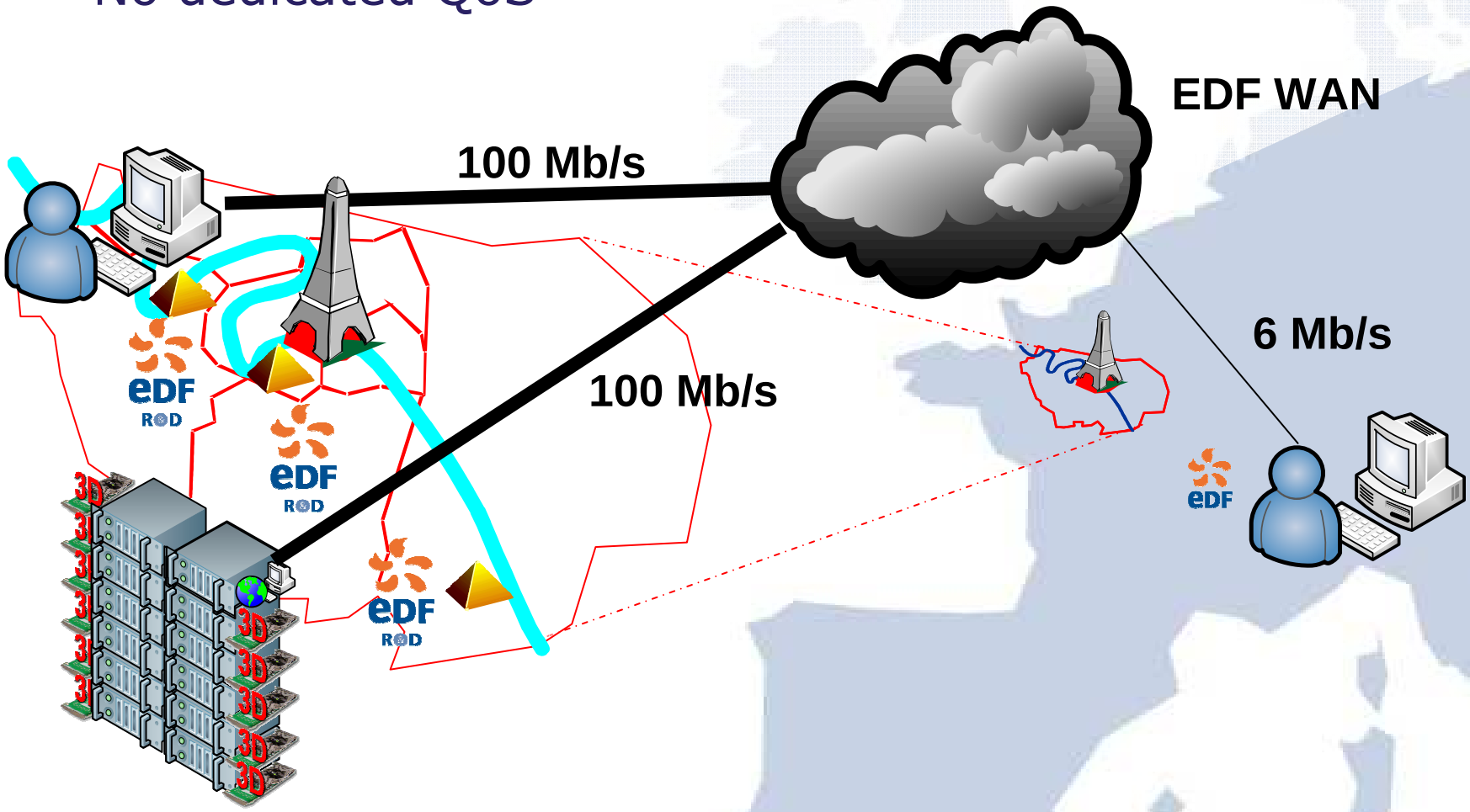
Experimental & ergonomic REX 1/7

- The first VISUPORTAL prototype was deployed on 11 graphics nodes cluster at EDF Clamart in October 2007.
- Tested between EDF R&D and an EDF engineering Unit (500 km far).



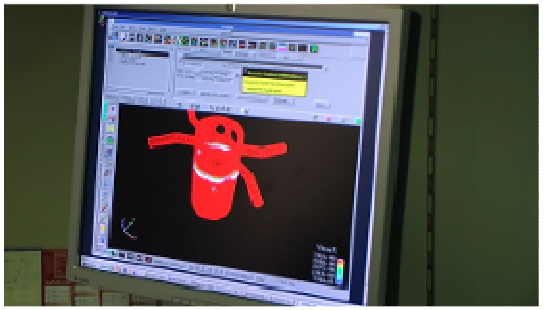
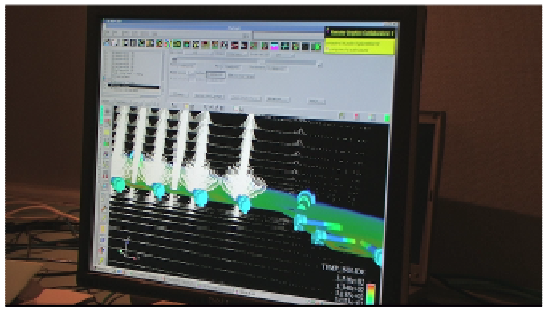
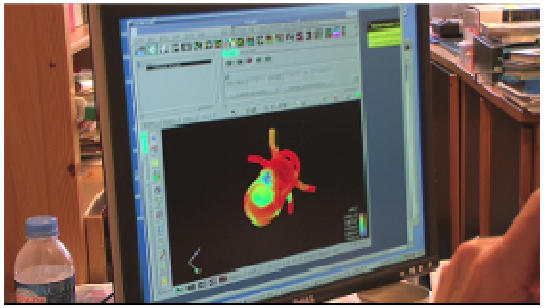
Experimental & ergonomic REX 2/7

- **EDF : A constrained WAN**
  - A narrow network bandwith with distant entities
  - No dedicated QoS

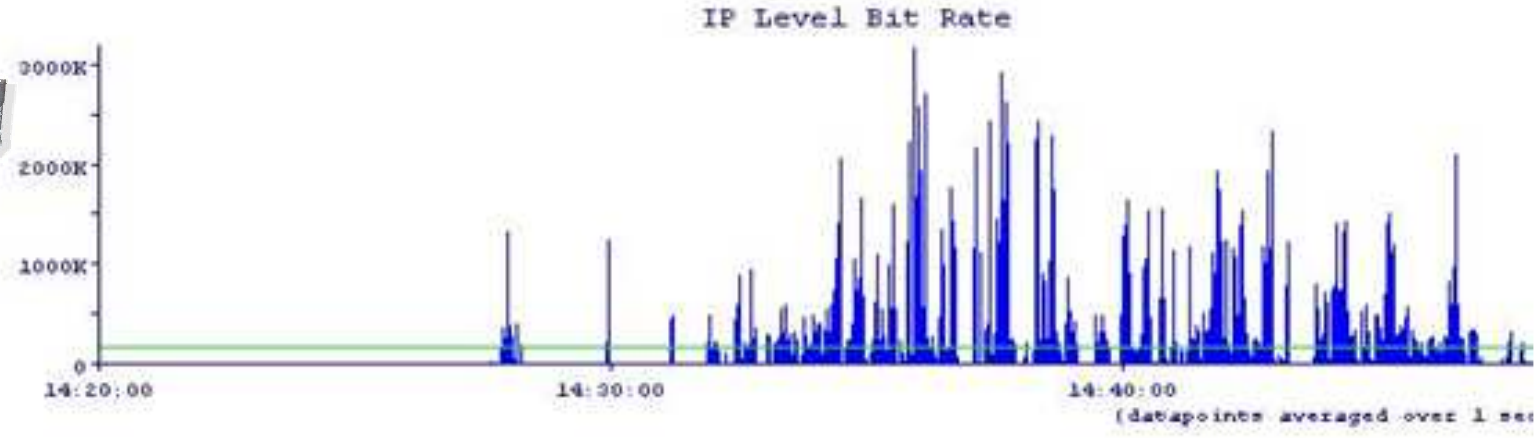


Experimental & ergonomic REX 3/7

- **3 pairs of distant users** :
  - Researcher (EDF R&D) /Engineer (EDF)
- On **real EDF case studies** (nuclear safety studies)

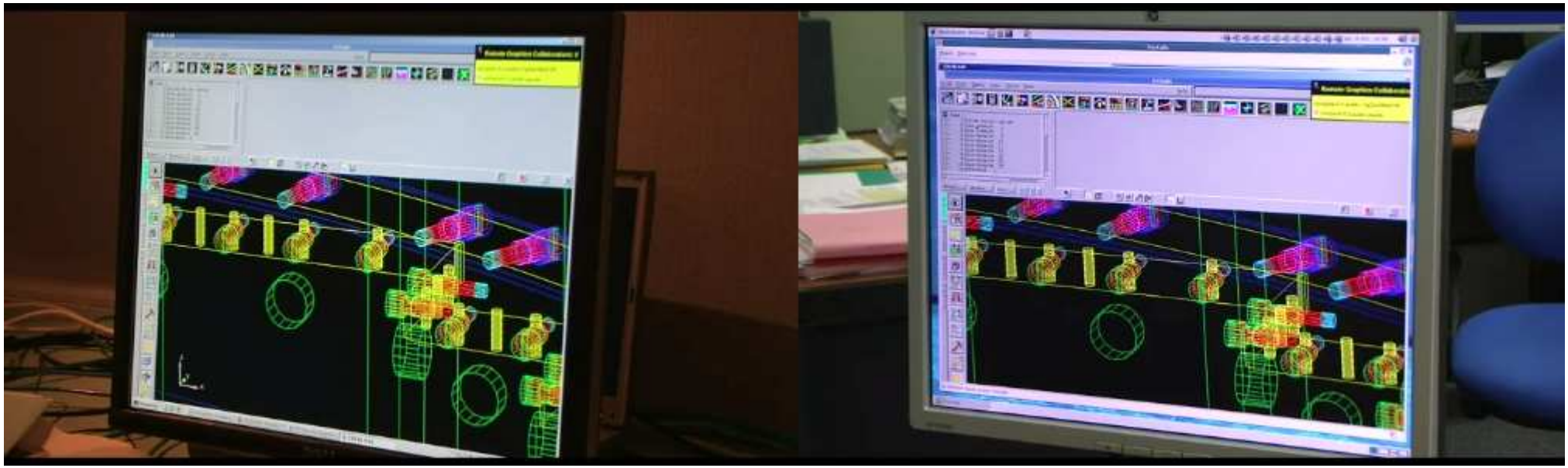




- Filmed and Networking measured/monitored on both sites
  - Two HD cameras,
  - Two Network traffic Analysers (Niksun NetVCR)



Experimental & ergonomic REX 4/7

- Short extracts of the video records of the experiment



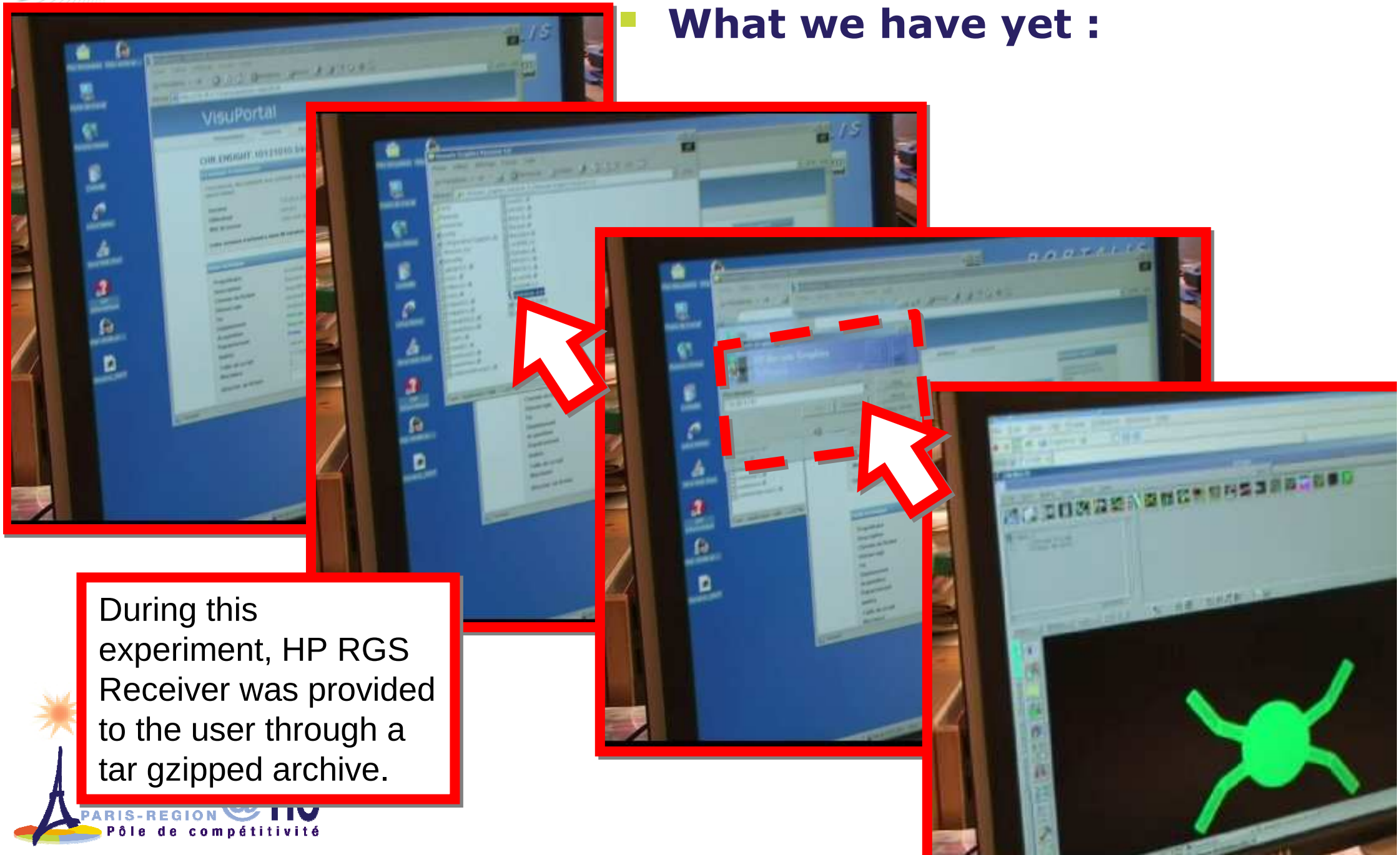
Experimental & ergonomic REX 5/7

- **« Is it possible to keep it ? »**
  - The « experimental »users are definitely convinced by
    - the « easy of use » of the Visuportal system
    - The cluster performance of EnSight software
    - the performance of HP RGS (even if **the maximum measured network bandwith was 2 Mb/s** (peak) for RGS)
  - The users never notice that they were using a distant graphic cluster.

- **But… a few disappointments :**
  - No easy-login system to connect the remote visualisation tool to the distant linux cluster nodes
  - No easy way to deploy the remote visualisation tool for new users.
  - Not enough intuitive Collaboration GUI

Experimental & ergonomic REX 6/7

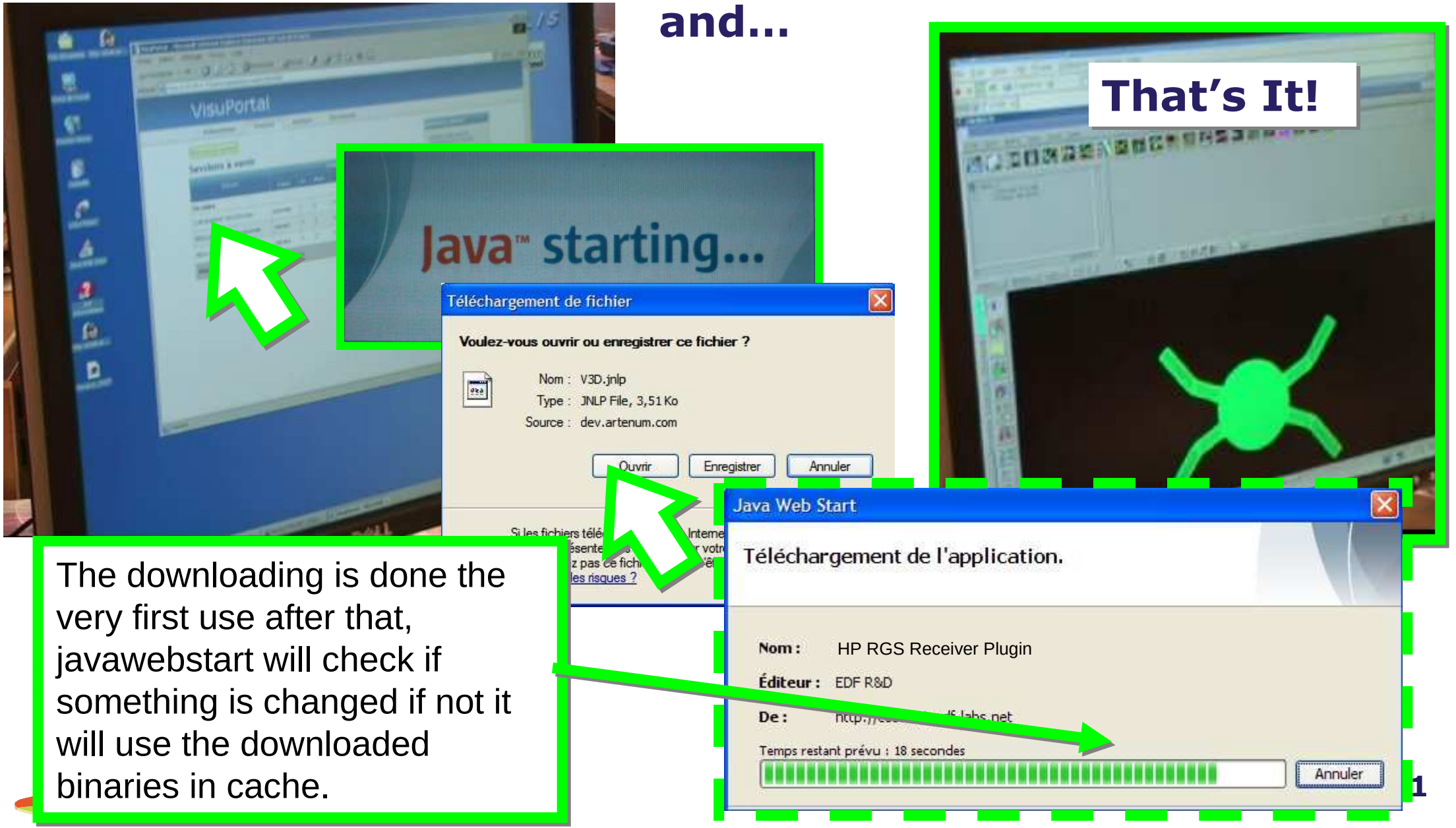
- **What we have yet :**



During this experiment, HP RGS Receiver was provided to the user through a tar gzipped archive.

Experimental & ergonomic REX 7/7

- **What the users want :just click on the session name and…**



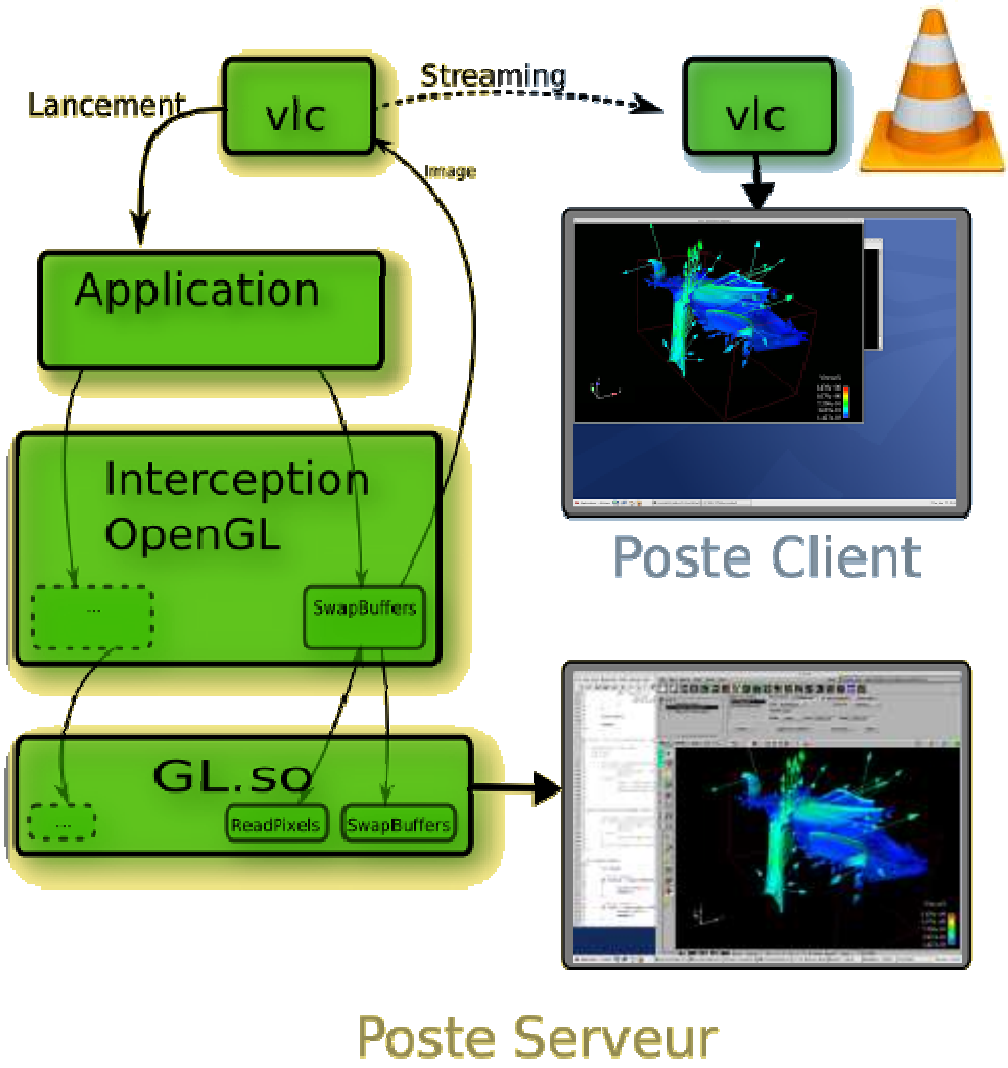
The downloading is done the very first use after that, javawebstart will check if something is changed if not it will use the downloaded binaries in cache.

Tests de performance de différentes solutions de visualisation distante

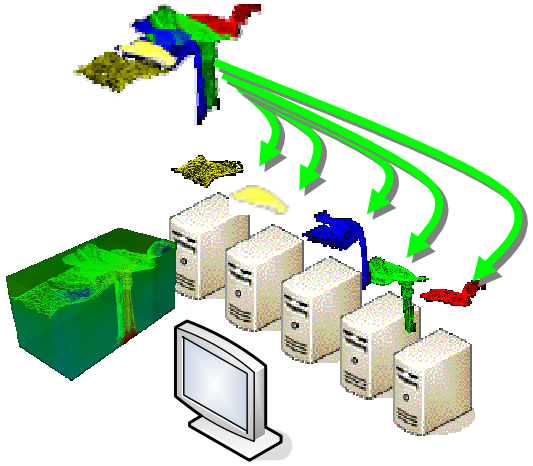
- Le projet a testé **la mise en œuvre, les performances** de solutions sur étagère matures et efficaces :
  - IBM DCV-RVN
  - HP Remote Graphics (HP RGS)

  Et testera les nouveaux produits
  - SUN shared visualisation system
  - SGI Visual Supercomputing (SC 2007)

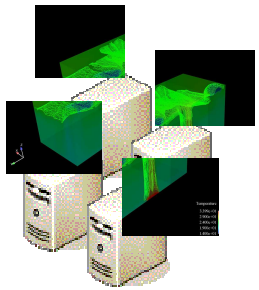
- VLC (Videolan Client) a été analysé, mis en œuvre et confronté à cet l'état de l'art :
  - La faisabilité a été démontrée avec le CEA
  - L'approche vidéo streaming nécessite un travail conséquent sur l'architecture de VLC pour atteindre les performances des solutions sur étagères.

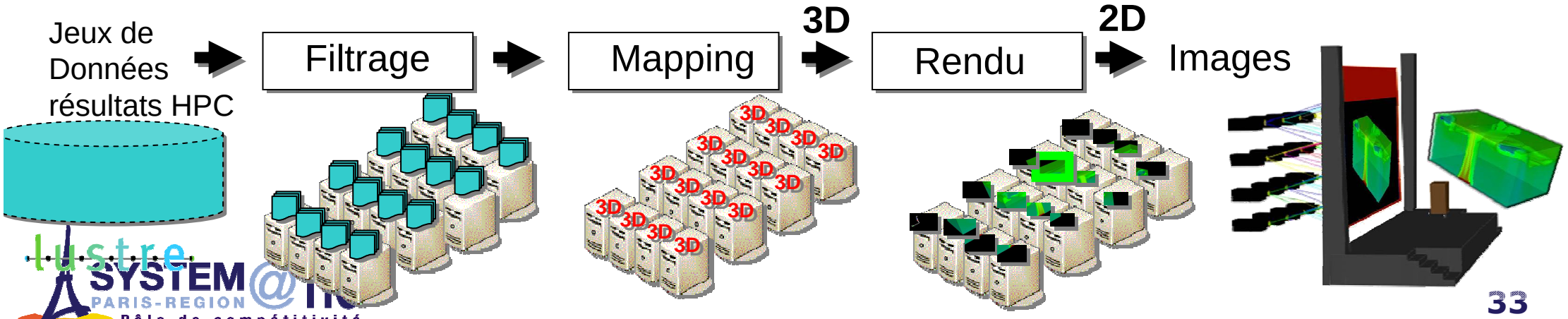


Scalability of visualization tools



- **Scalability : a complex issue**
  - Datasets size (more « points »)
  - Display size (more pixels)

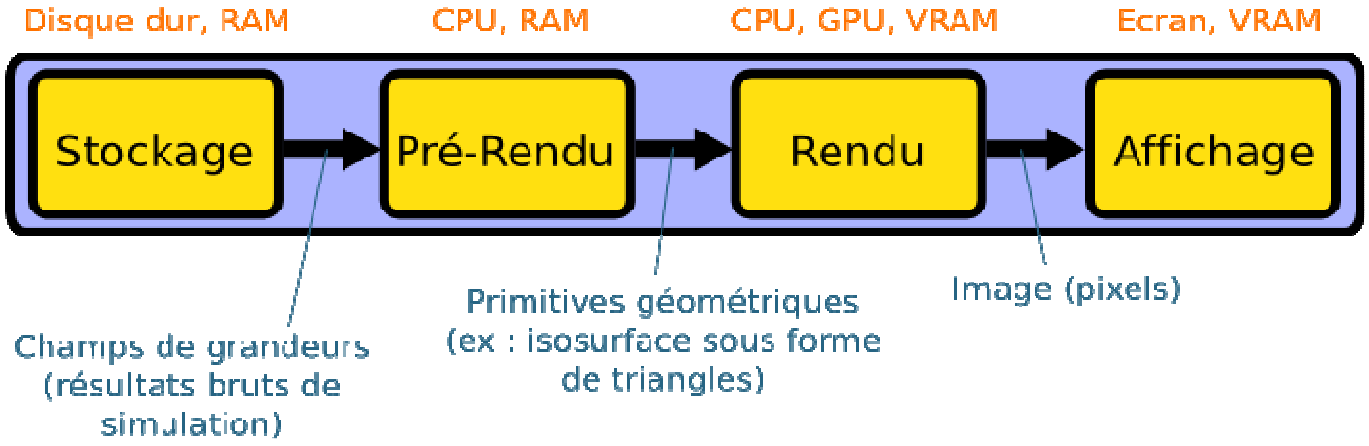




Back-of-the-envelope analysis

- 40 Gb/s : how to consume bandwidth ?
- Straightforward approach of image transmission, no compression
  - 32 bits, stereo (x2)
  - 15 fps
  - Max image size of 40 G / 64 / 15 ~ 40 Mpixels…
    if 100% efficient – and data access/mapping/rendering is fluent !
  - MIRAGE CEA display is already 14 Mpixels
  - Of course compression can be used of other levels of transmission:
    - Data
    - Geometry

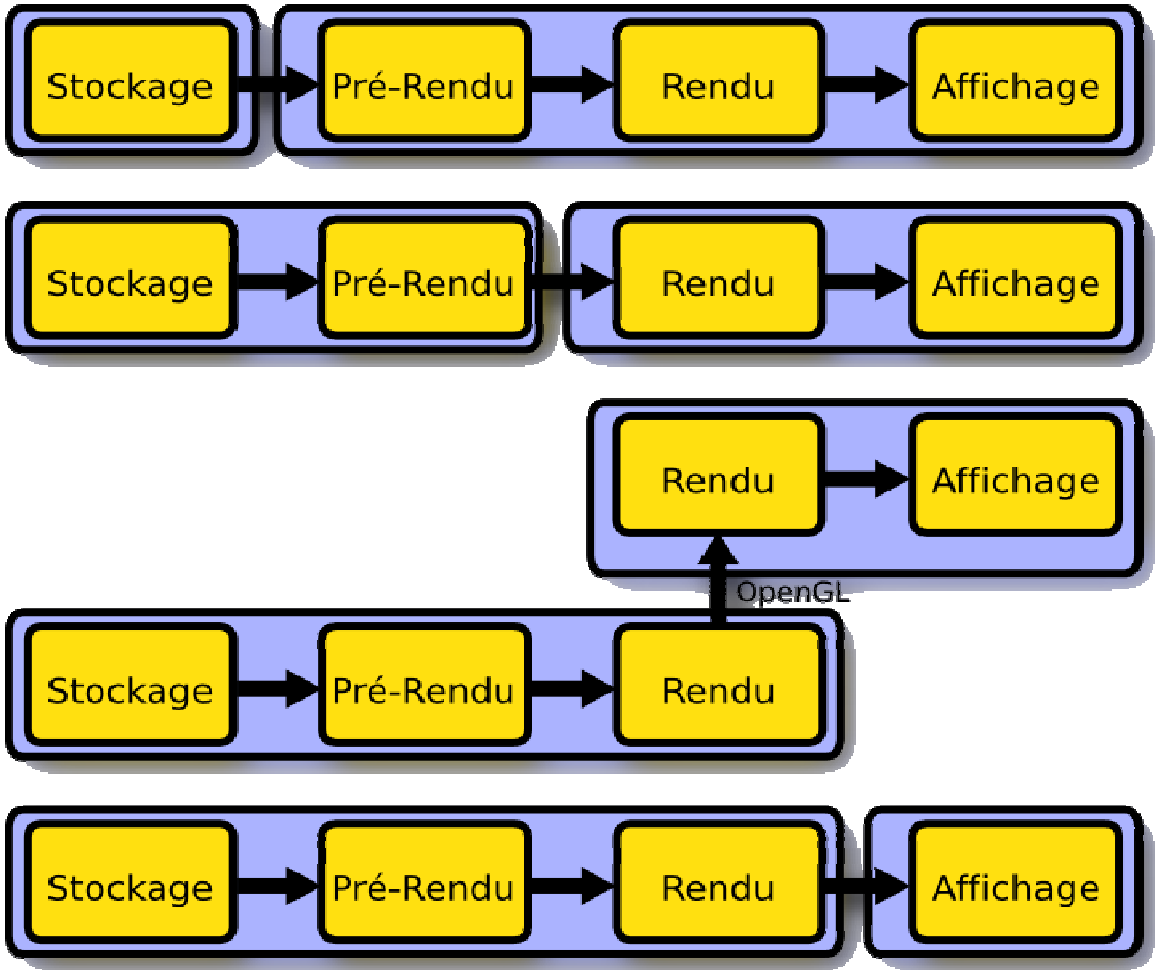
  splitting the visualization pipeline at different stages



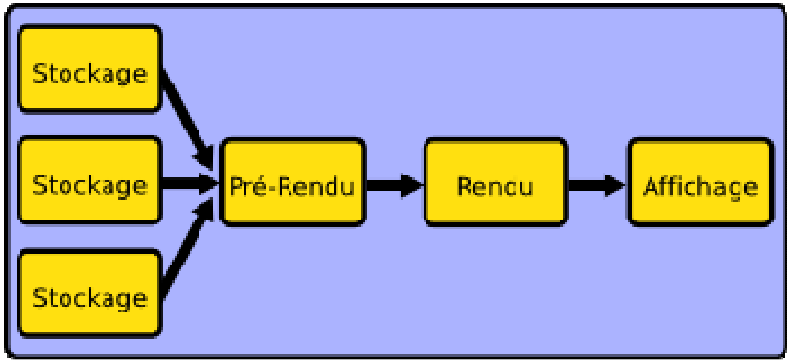
Large scale visualization

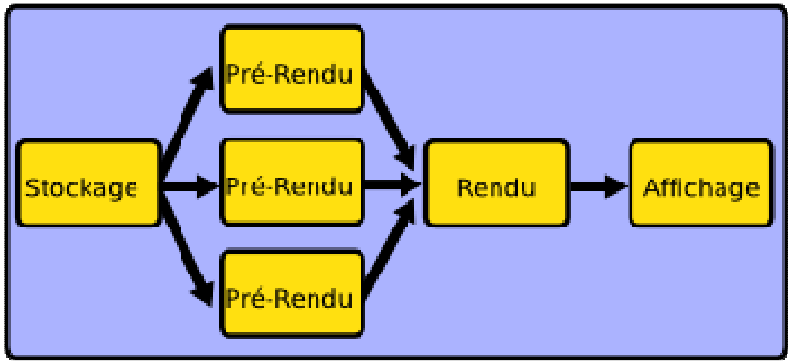
- 2 driving parameters
  - Datasets size
  - Display size
  - … and an output parameter = ‘fps’
- Parallelism is a common answer to deal with these complexity factors
  - Datasets => more “data management processes”
  - Pixels => more “ rendering processes”
  - Not the ultimate approach (you can be smarter than brute force “divide and conquer”) but it is quite straightforward and general-purpose enough

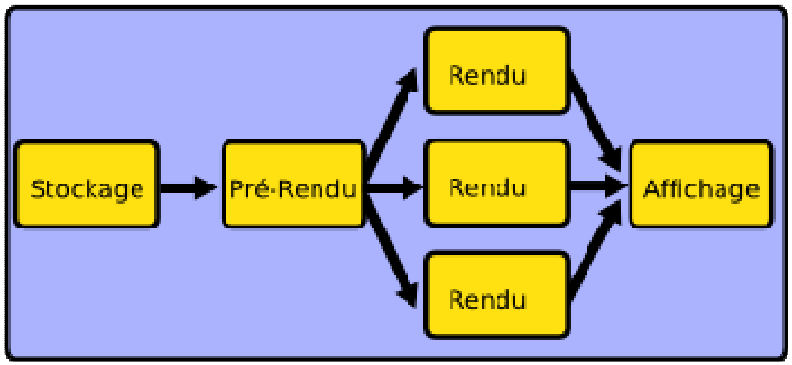
Viz pipeline possible distribution

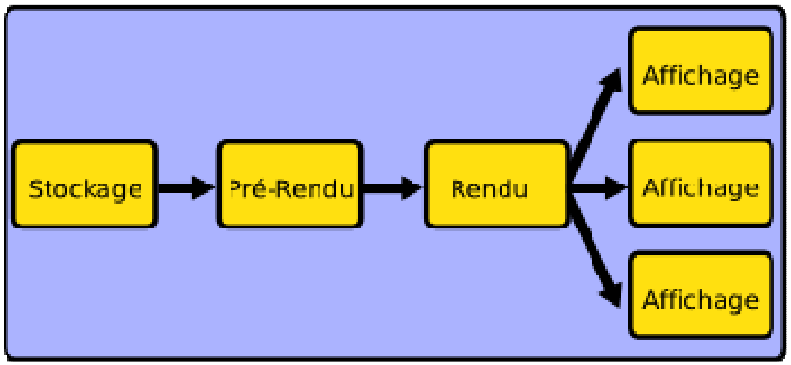


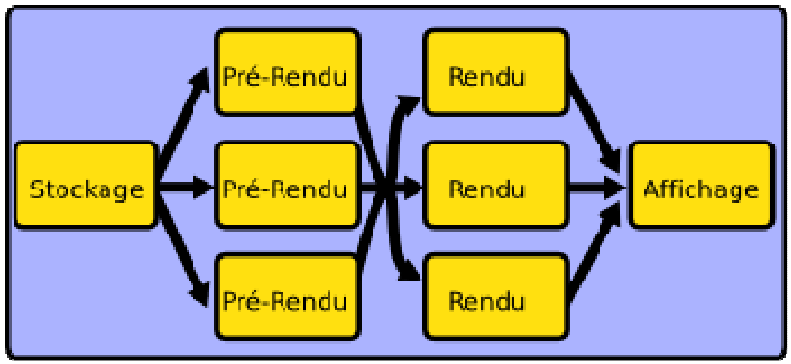
Each stage can be “locally” parallelized





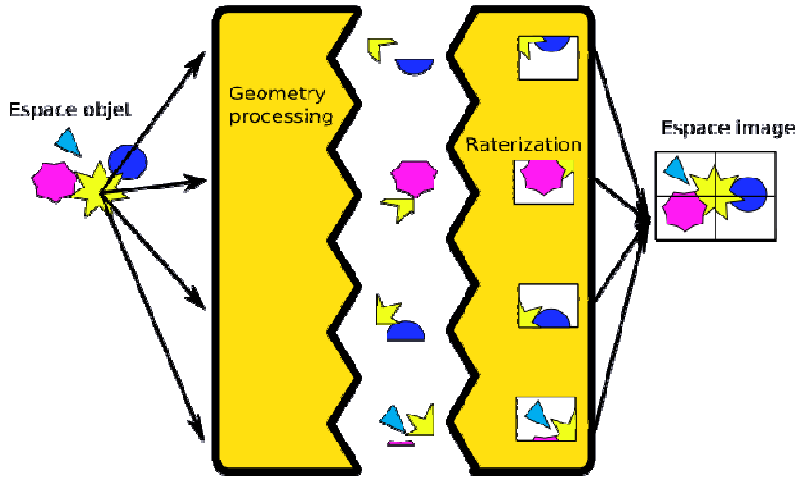






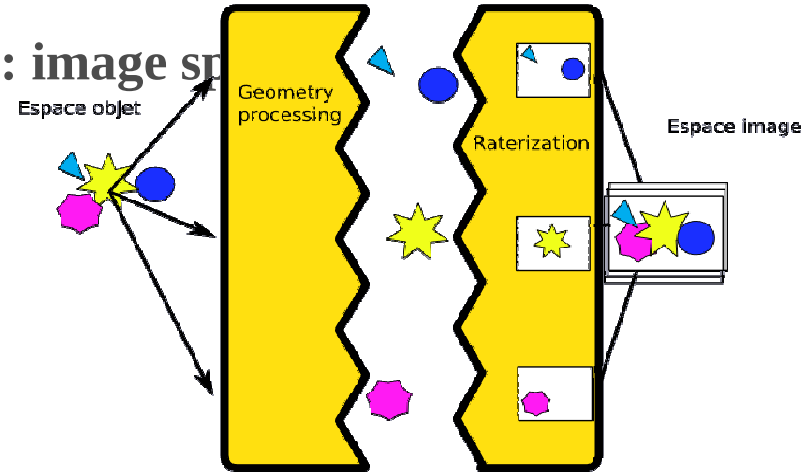
Parallel rendering reminder / compositing

- **Sort First : object space**



Scalable w.r.t. resolution

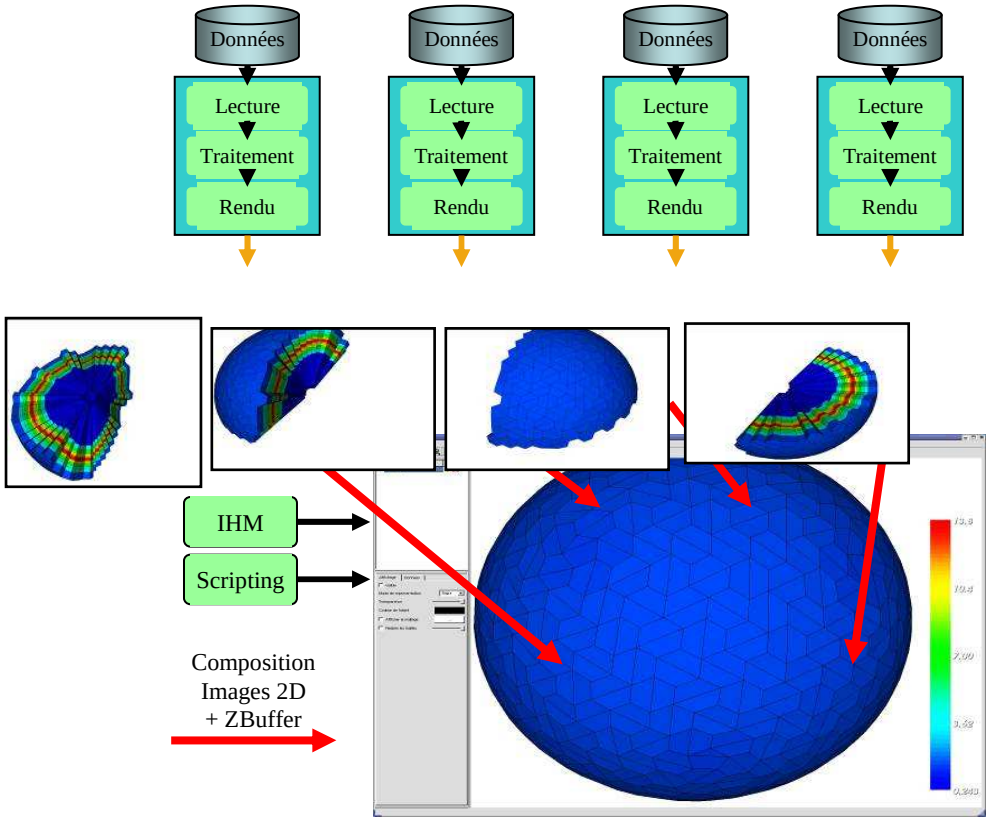
- **Sort Last : image sp**



Scalable w.r.t. dataset (geom) size

CEA/DIF experience

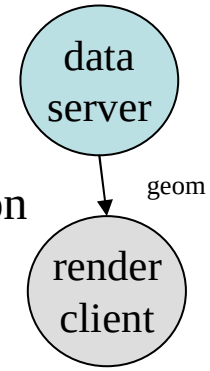
LUSTRE everywhere (TERA)…
VTK, EnSight
On clusters and large displays
“Weakly remote”

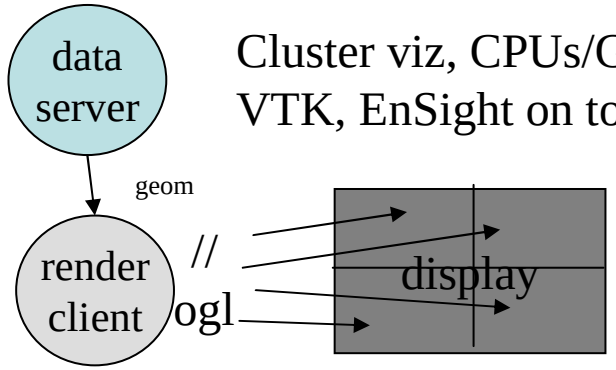


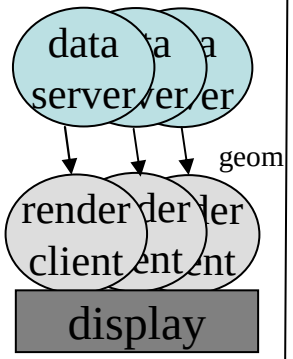
rendering (client) parallelism

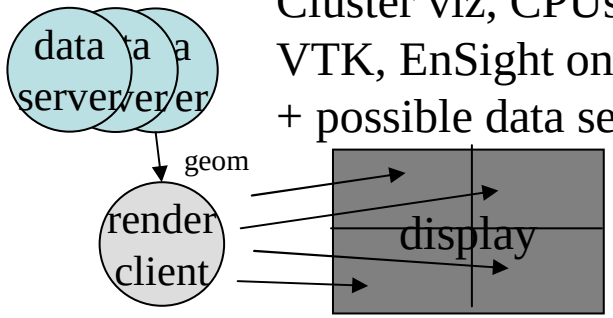
Data (server) parallelism

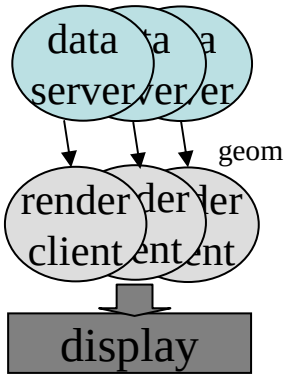
	Display = desktop	Display = large display
Small size Datasets <5M cells	Local viz one CPU/GPU + remote server option VTK, EnSight	Cluster viz, CPUs/GPUs VTK, EnSight on top of Techviz
Medium size Datasets <100 Mcells	Cluster viz+rendering Desktop delivery VTK (EnSight…)	Cluster viz, CPUs/GPUs VTK, EnSight on top of Techviz + possible data servers in //
Large Datasets >100 Mcells	VTK on TERA s/w rendering desktop delivery	EnSight DR VTK + ICE-T (work in progress)

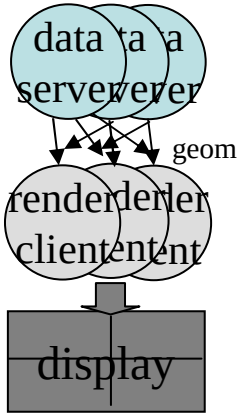




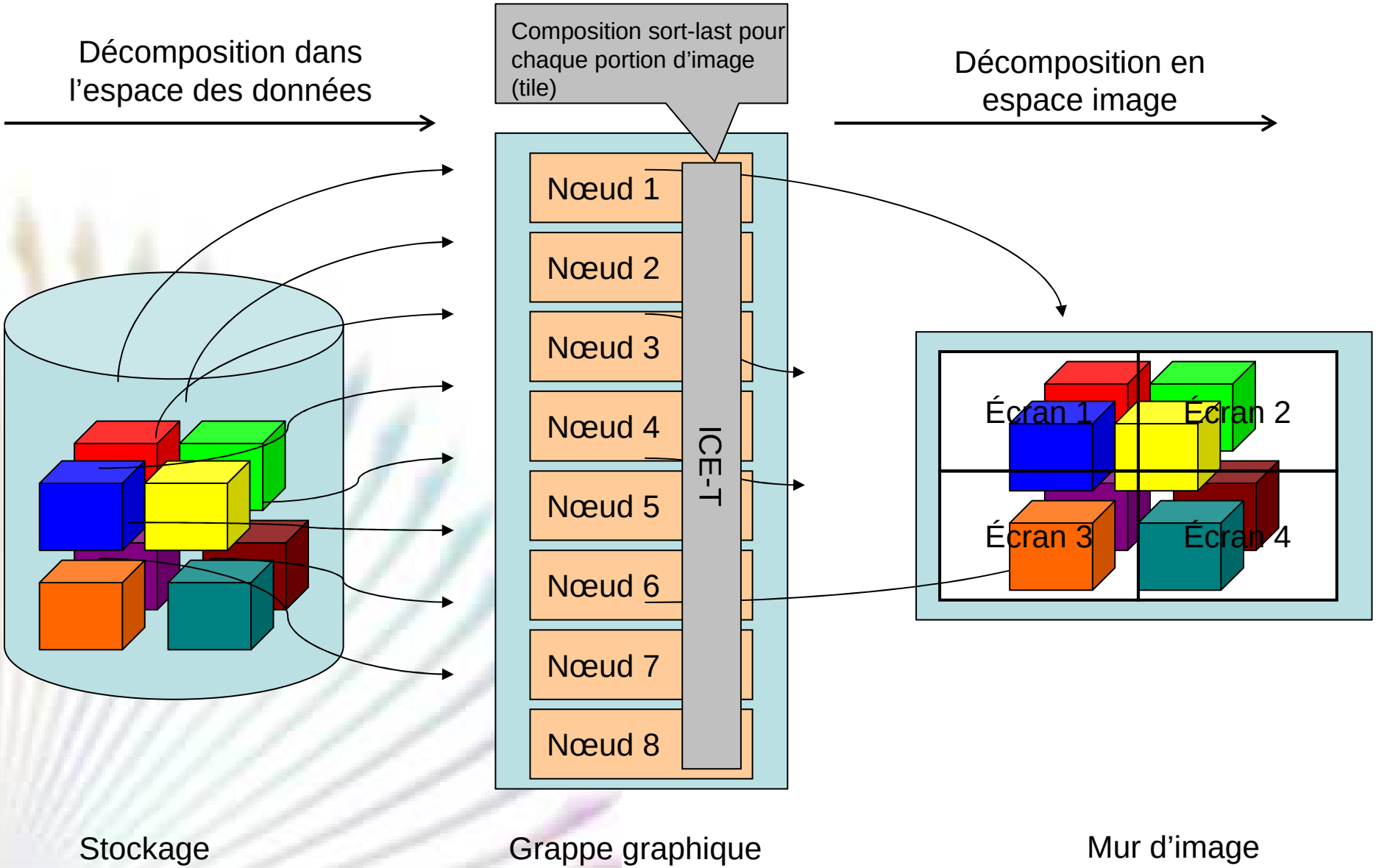








Hydrid SF/SL : ICET-T limited to VTK (pull data-flow model)



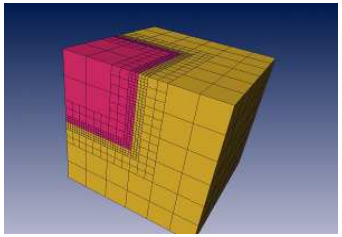
Overall approach – WP4

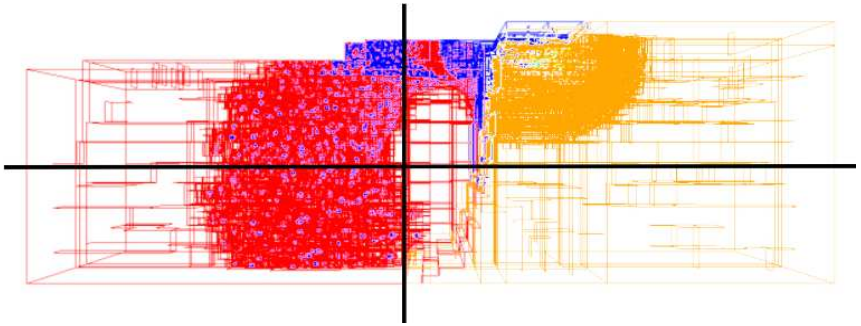
- Limited manpower and strong concern of perennity and independence
- Use:
  - existing software standards
  - open source components

  as much as possible
- This is not in contradiction with the usage of proprietary software
  - as long as it can be mixed with open components

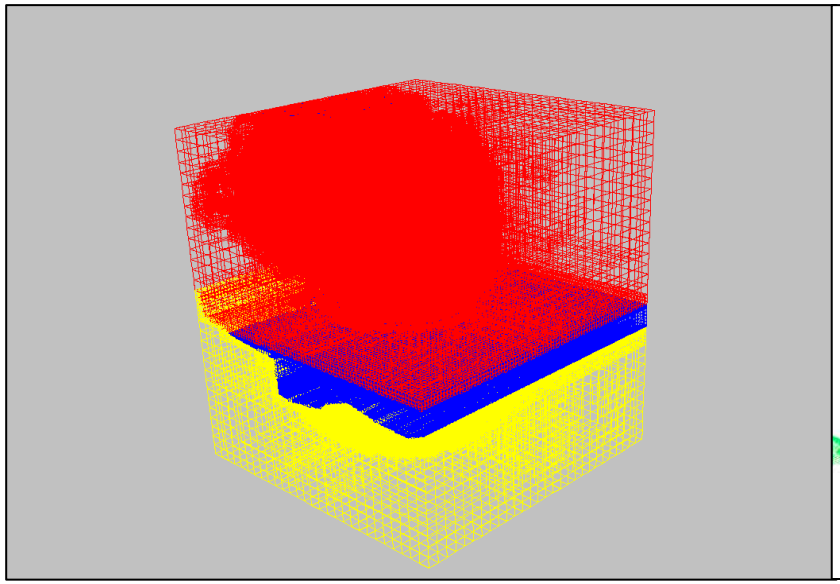
- 2007 = benchmarks, prototypes on small cluster/tiled display
- 2008 = consolidate, connect with LUSTRE and network
- 2009 = deploy and evaluate

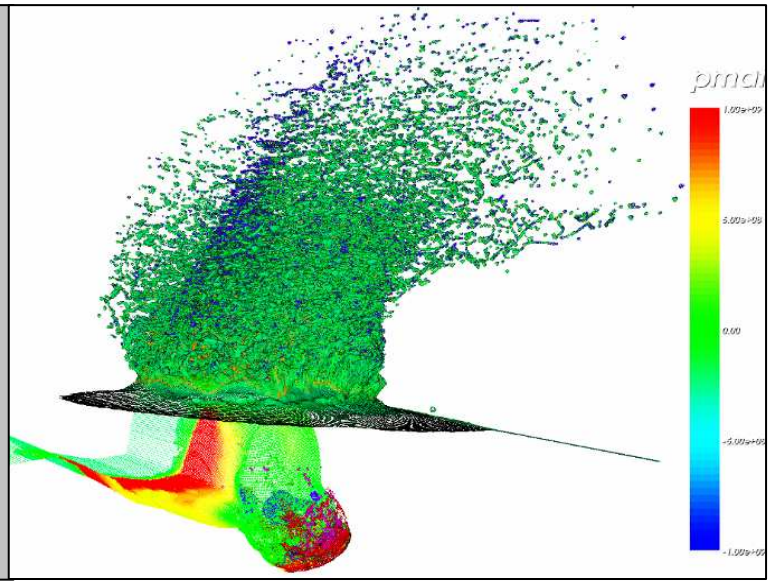
CEA "use case" (basic benchmark for parallel viz)



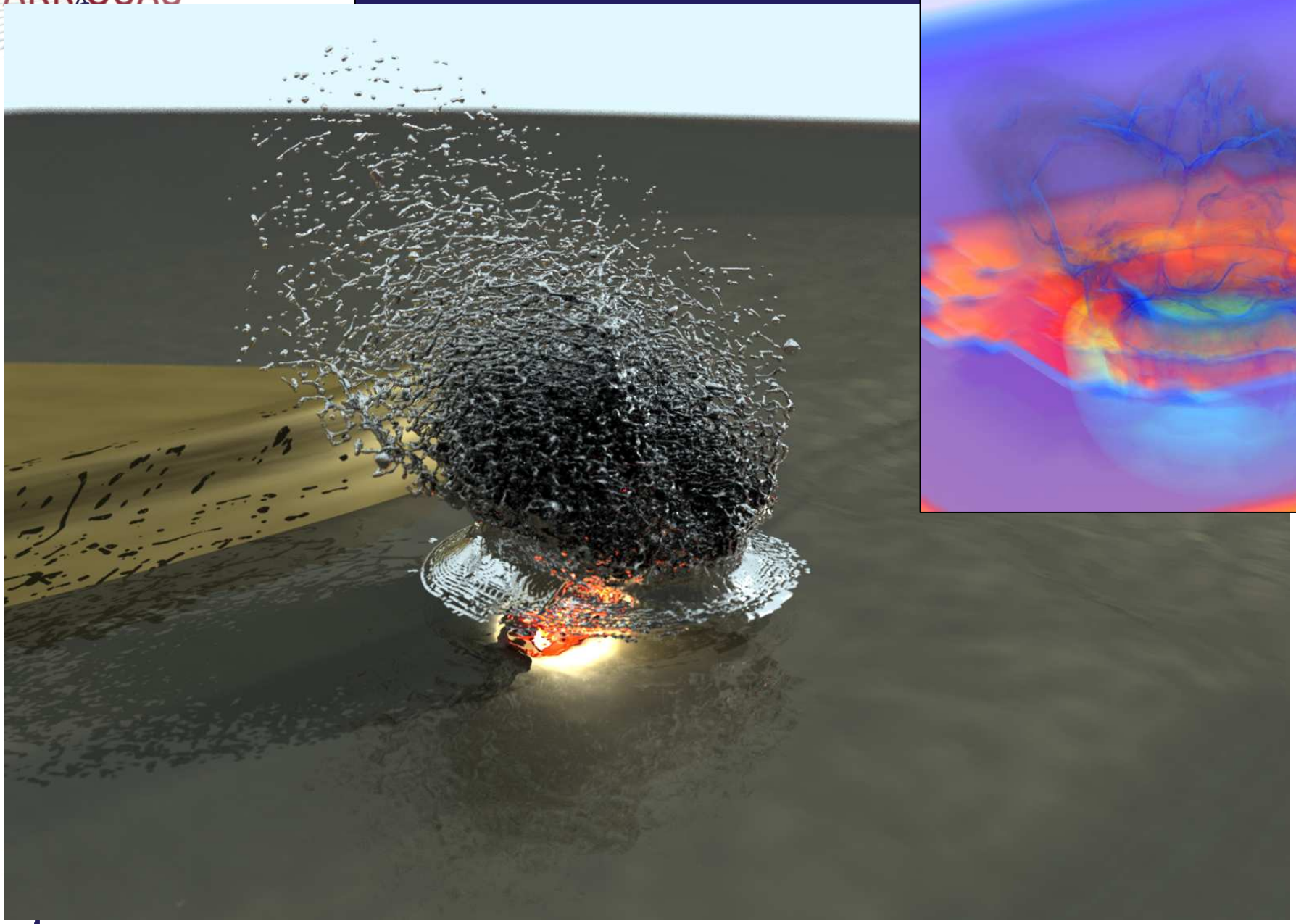


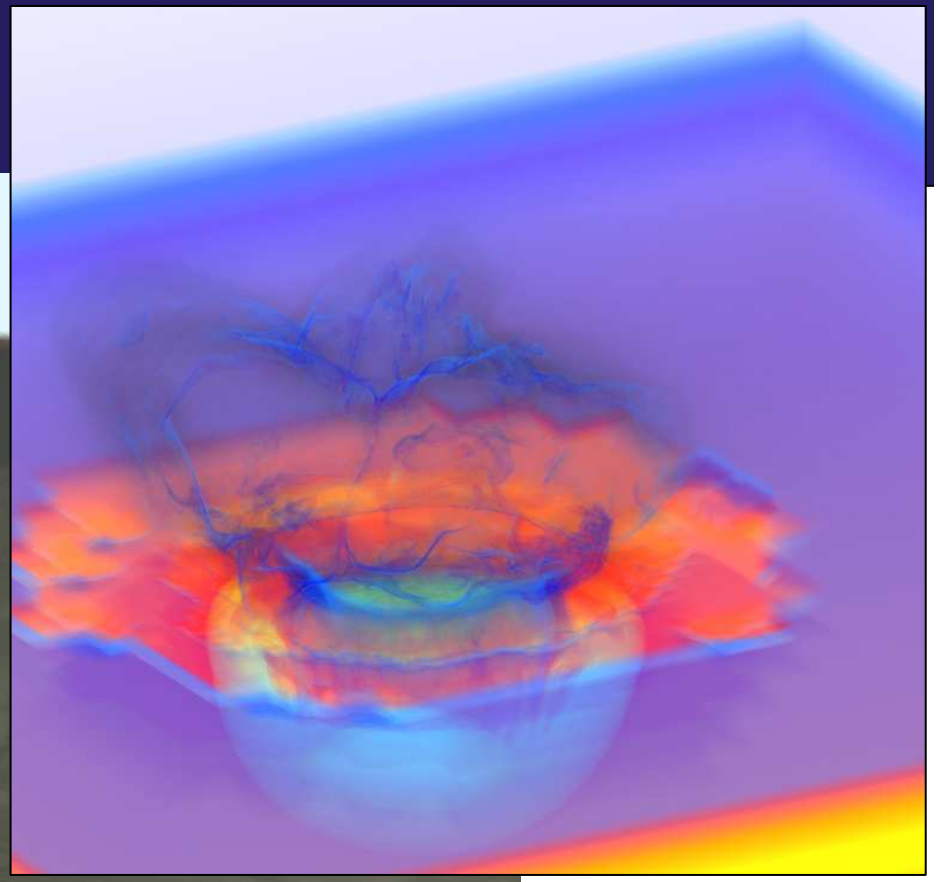
- A typical CEA parallel simulation : "Meteor" dataset
  - AMR, cartesian – considered as non structured (generic CEA concern)
  - 15 to 120 Mcells
  - 64 to 96 domains (mesh decomposition)





CARROCAS





SYSTEM@TIC PARIS-REGION Pôle de compétitivité

Very pragmatic focus and choices

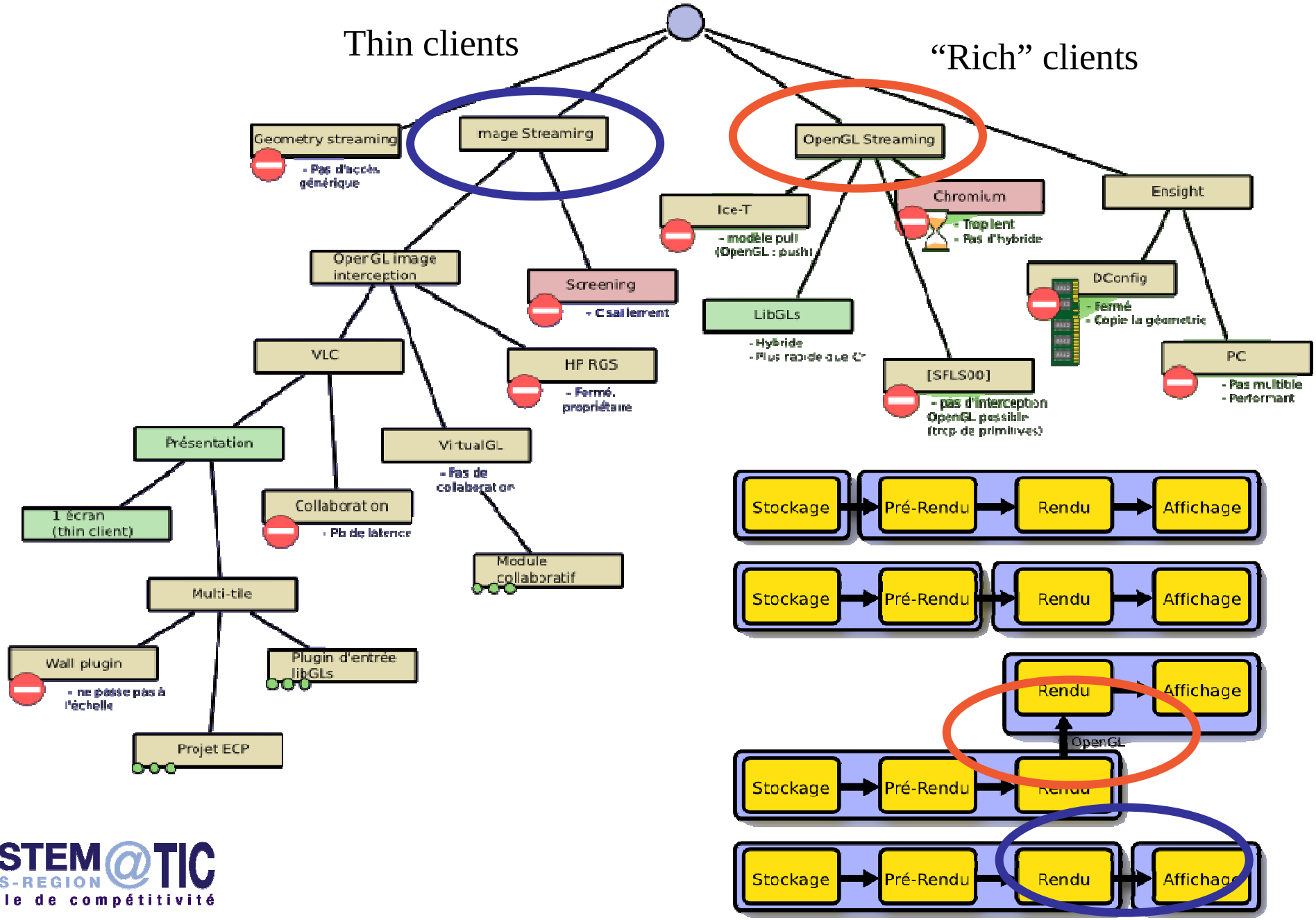
- General purpose existing packages, several “fairly good ones”
  - VTK / PARAVIEW
  - EnSight
  - AMIRA
  - VisIt…
  - Quite strong intersection but not strictly equivalent

- Focus on common denominator already used at CEA and EDF
  - EnSight
  - And VTK as a common kernel to PARAVIEW, CEA own developments - and possibly VisIt

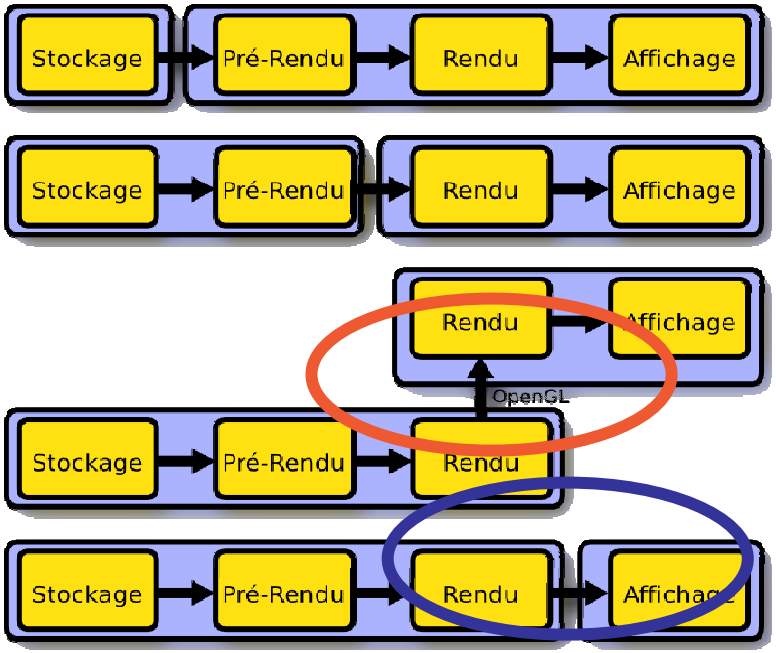
What about "middleware"?

- Lower level layers able to parallelize / distribute rendering
- Either compositing, OpenGL interception or image streaming layers
  - OGL
    - Cr (Chromium)
    - Techviz (widely used in "local" production on CEA MIRAGE display)
    - …
  - Compositing
    - DVIZ
    - ScaleViz
    - Equalizer
    - …
  - Image streaming
    - OpenGL VizServer
    - IBM DVC
    - HP RGS
    - VLC ?
  - Only Cr and Equalizer (and VLC) are open source
  - Cr comes bundled with EnSight DR => try it…
  - We focused on this level of OGL or image remote articulation (genericity / transparency)

CEA experiments…



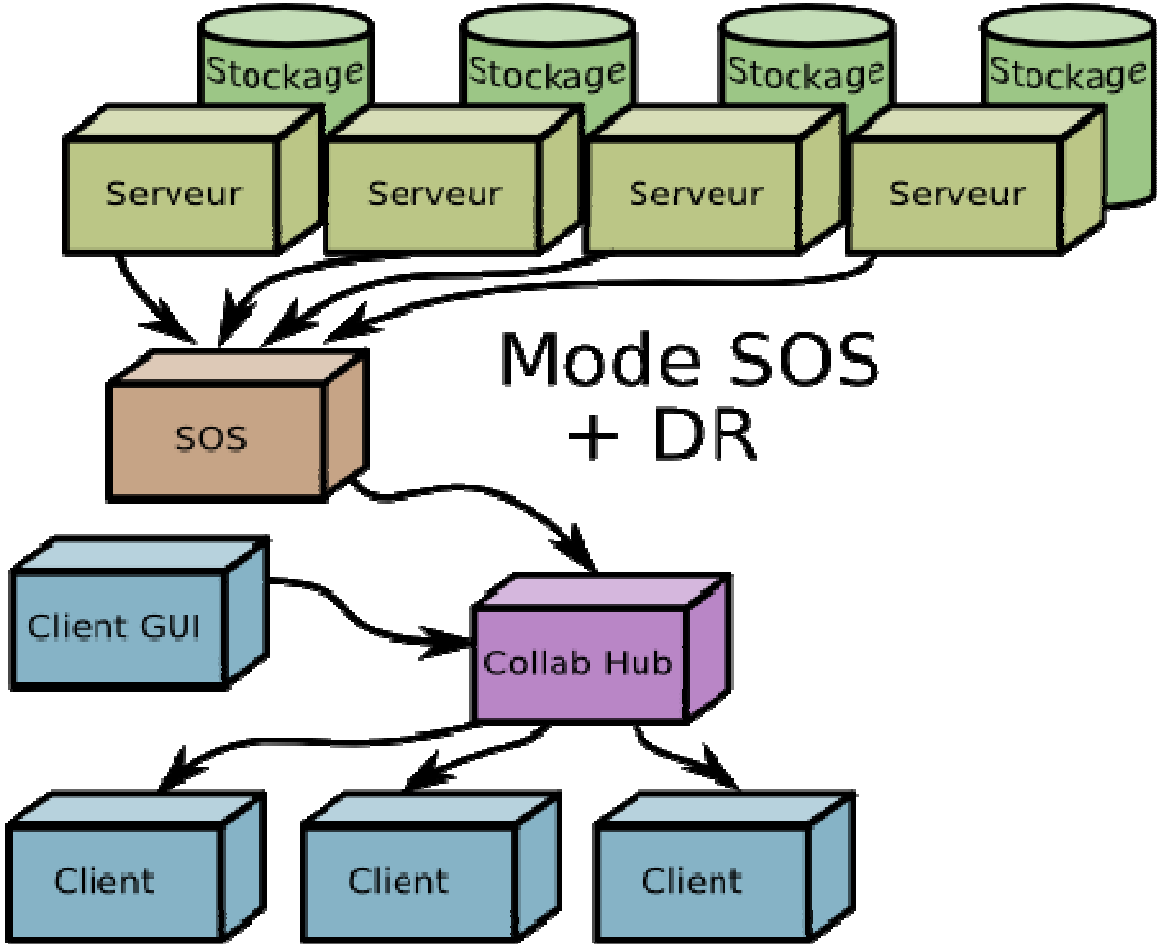
- Parallel data access and pre-rendering
  - In charge of the application
- Generic approaches, application independent
  - And a good location to consume bandwidth…



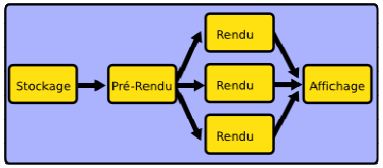
Preliminary setup : parallel visualization

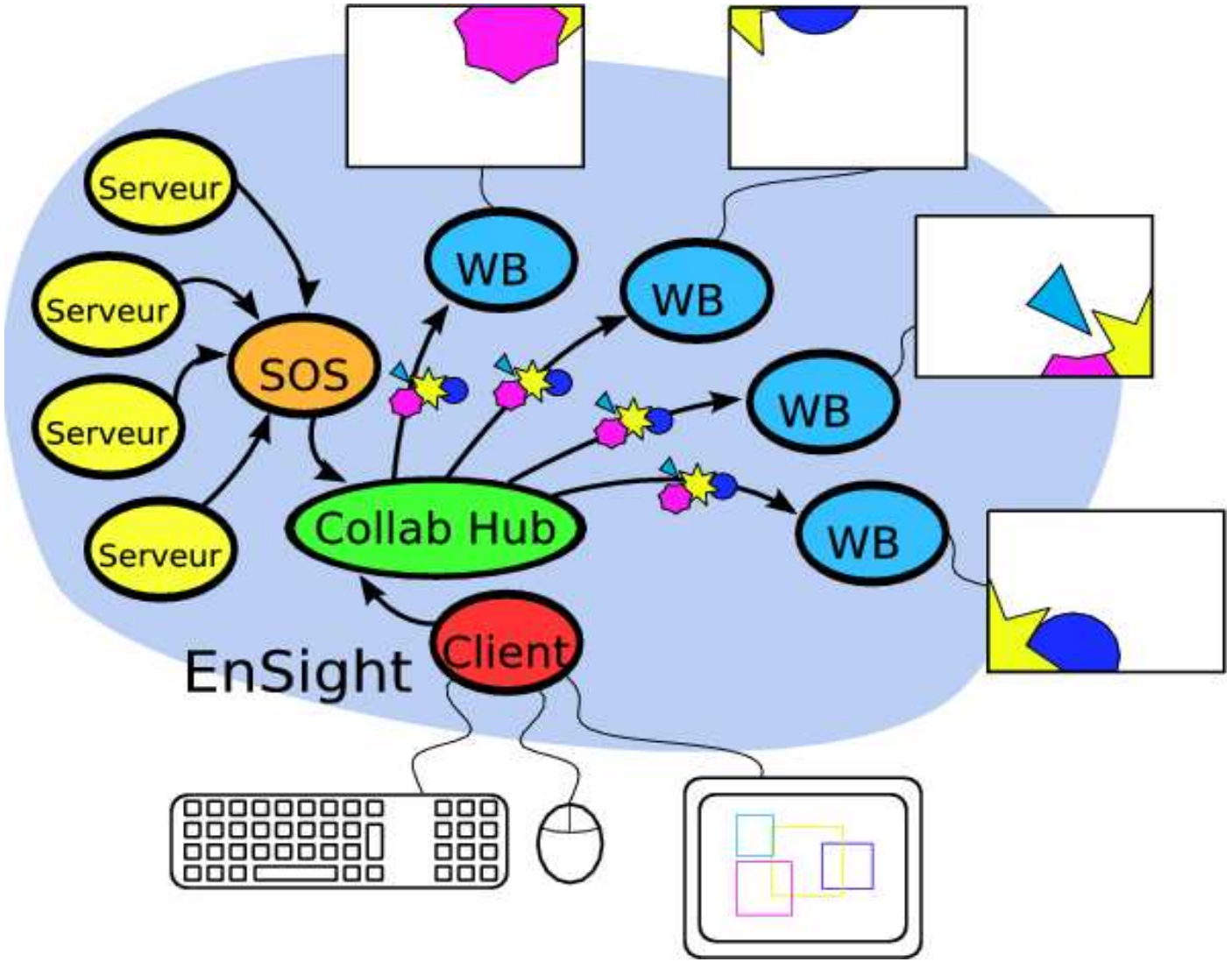
- EnSight DR evaluation : with support of DISTENE at TER@TEC – on TER@TEC visualization reseach facility

+ 8 node “small” graphics cluster



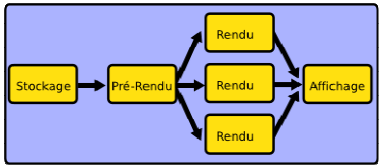
Parallel rendering, DCONFIG mode

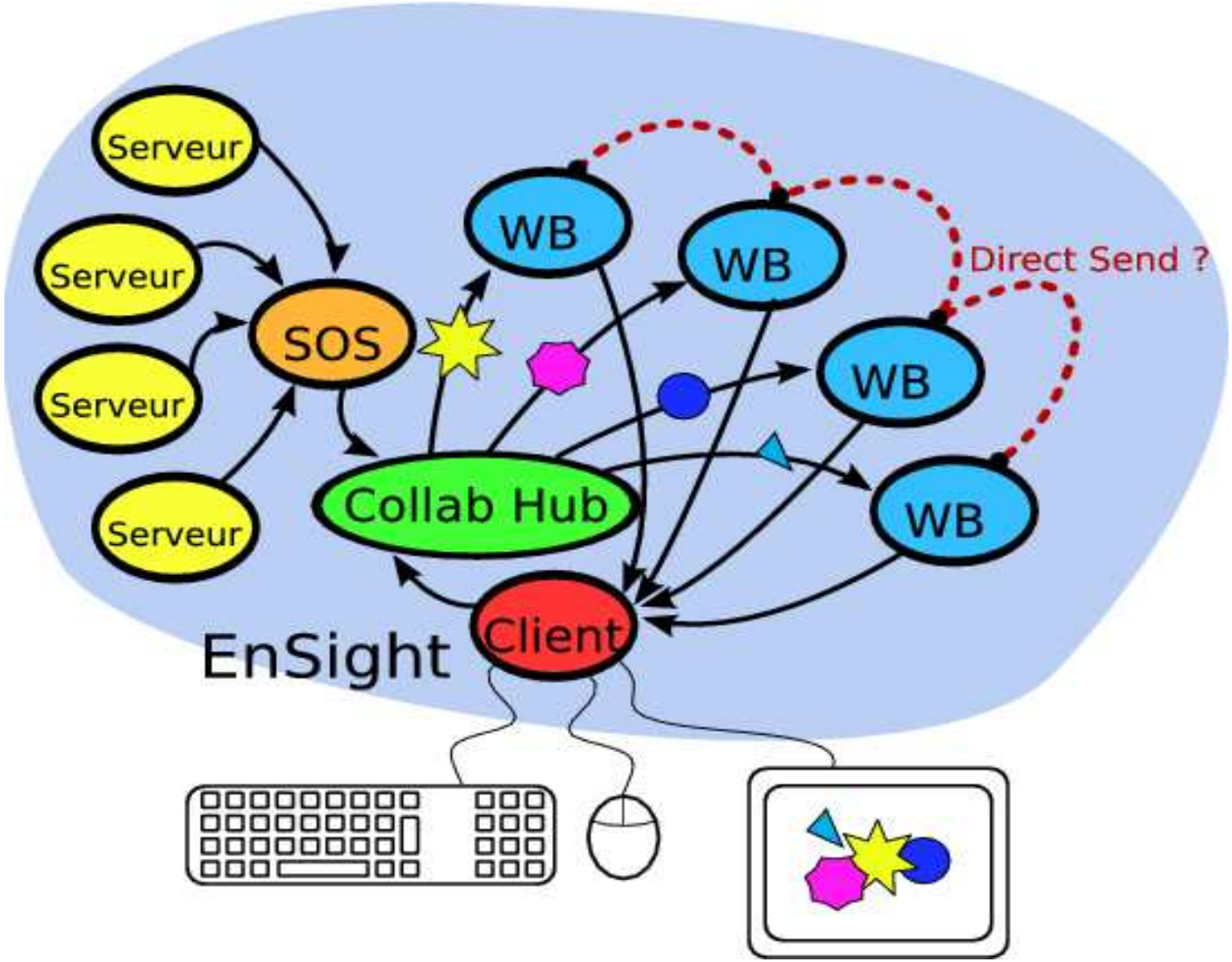




- Géométrie dupliquée sur tous les noeuds WB
- Sort-First (passe à l'échelle en résolution)
- Occupation mémoire totale linéaire en fonction du nombre d'écrans (tiles)
- Temps de rendu limité par la capacité de rendu d'une machine

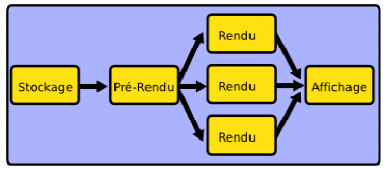
Parallel rendering, PC mode

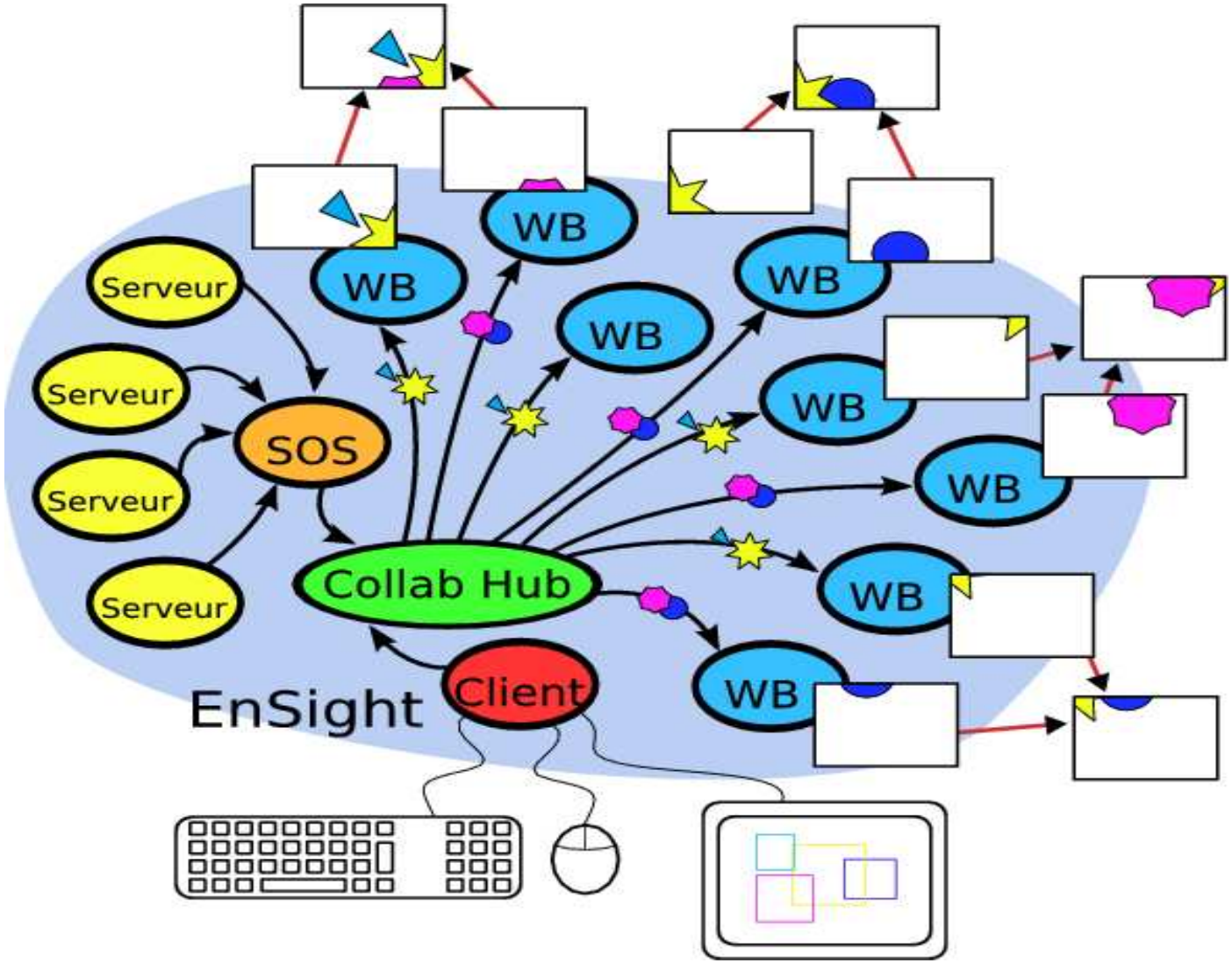




- Géométrie distribuée sur tous les noeuds WB
- Sort-Last (passe à l'échelle en complexité)
- Occupation mémoire totale fixe (la même que sur 1 seule machine)
- Temps de rendu limité par le réseau

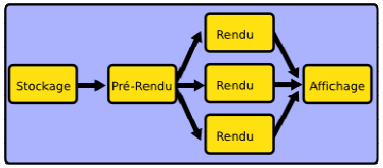
Parallel rendering, DCONFIG WH mode

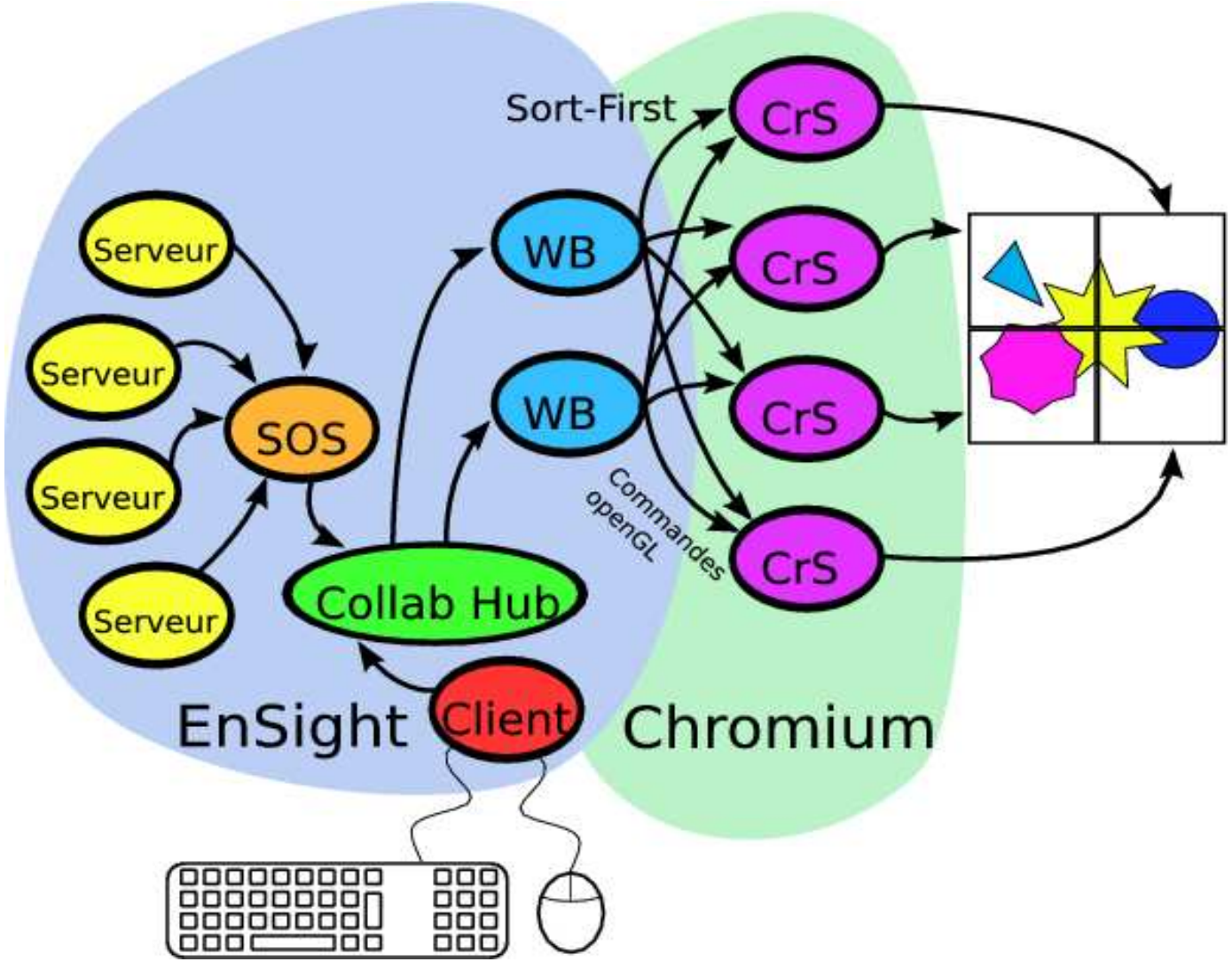




- Géométrie dupliquée sur tous les groupes de noeuds WB (distribuée au sein d'un groupe)
- Sort-First + Last (passe à l'échelle en complexité et en résolution)
- Occupation mémoire totale linéaire en fonction du nombre d'écrans (tiles)
- Temps de rendu limité par le réseau

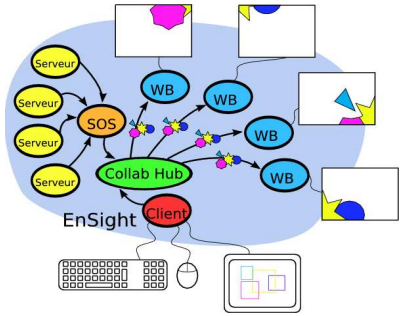
Parallel rendering, Chromium mode

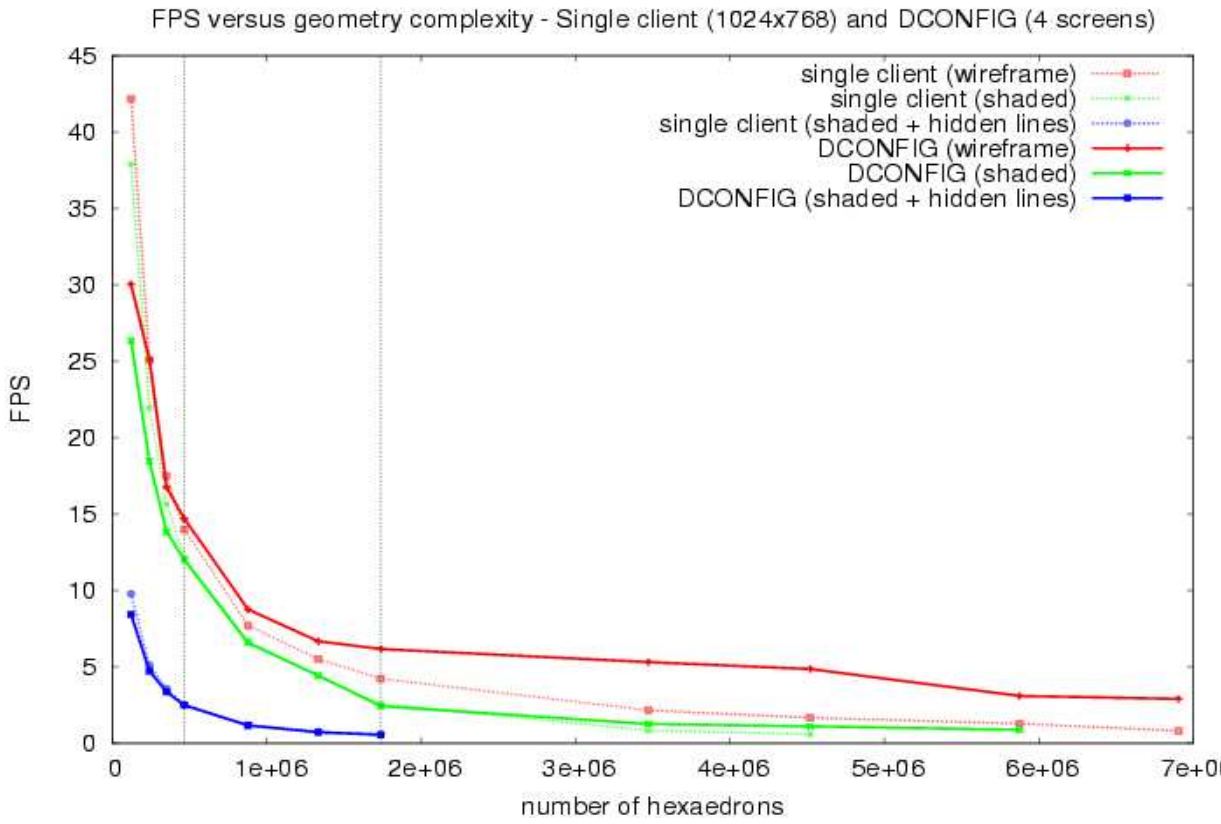


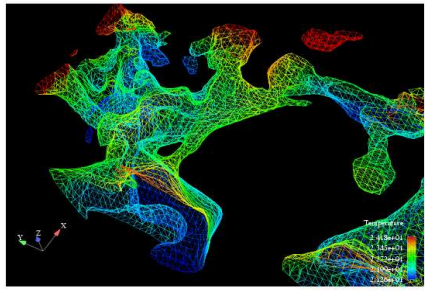


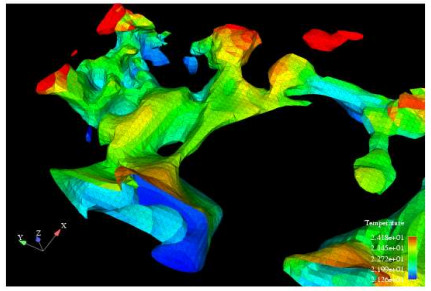
- Géométrie distribuée
- Sort-First (passe à l'échelle en résolution)
- Occupation mémoire totale constante
- Temps de rendu limité par la capacité de rendu d'une machine et par le réseau

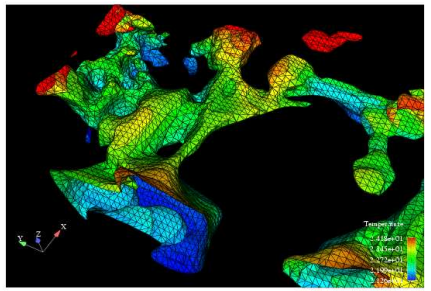
Benchmarks, DCONFIG



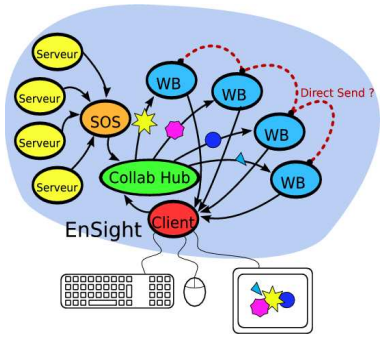


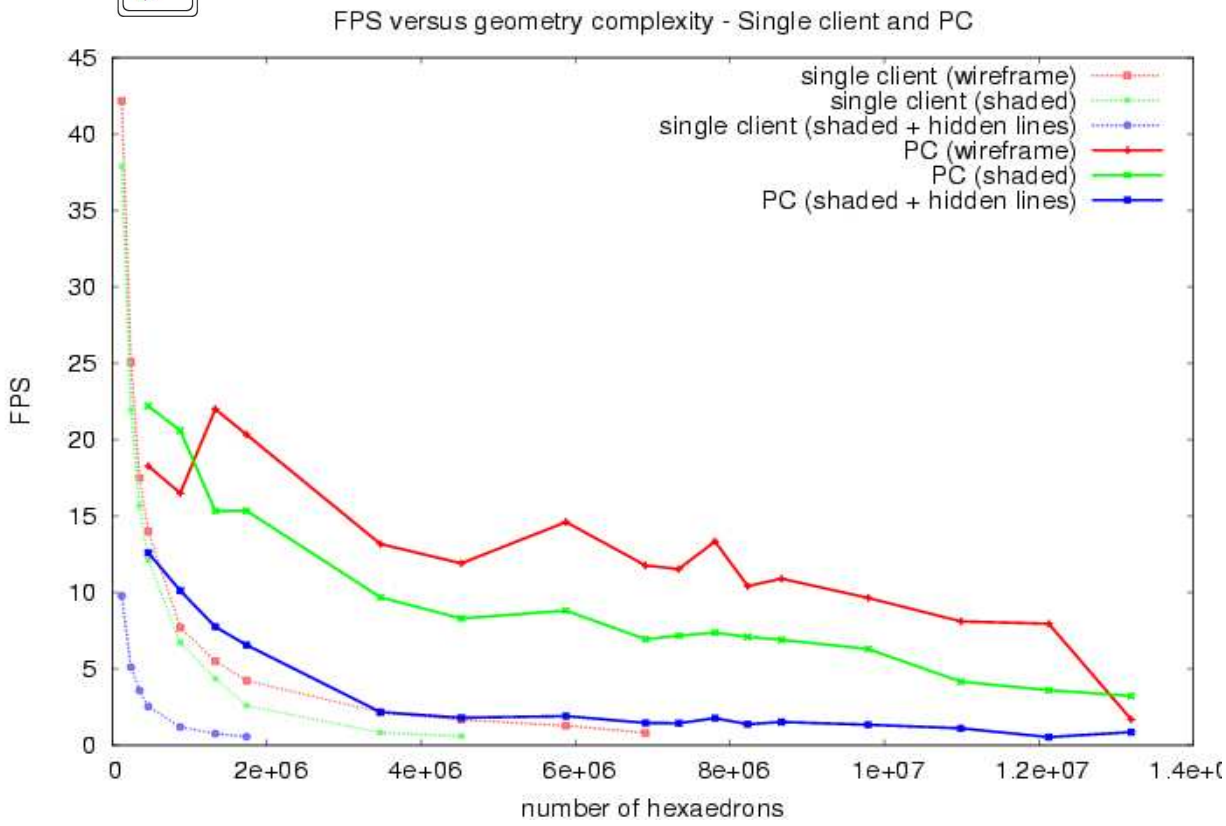


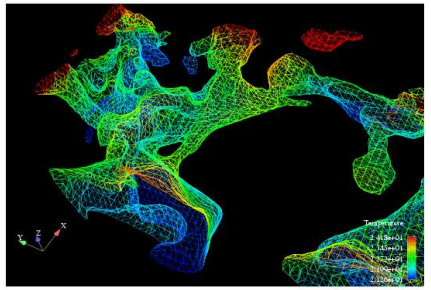


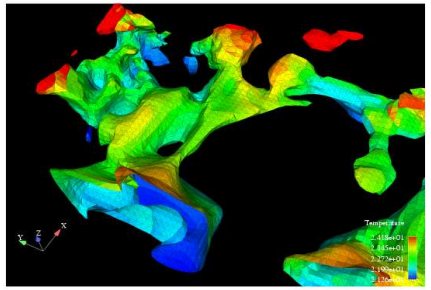


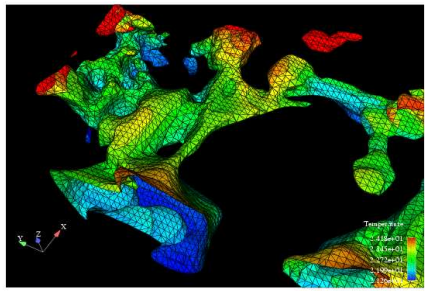
Benchmarks, PC



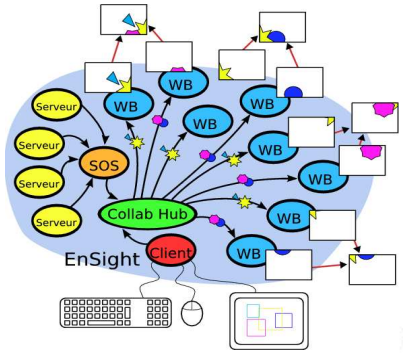


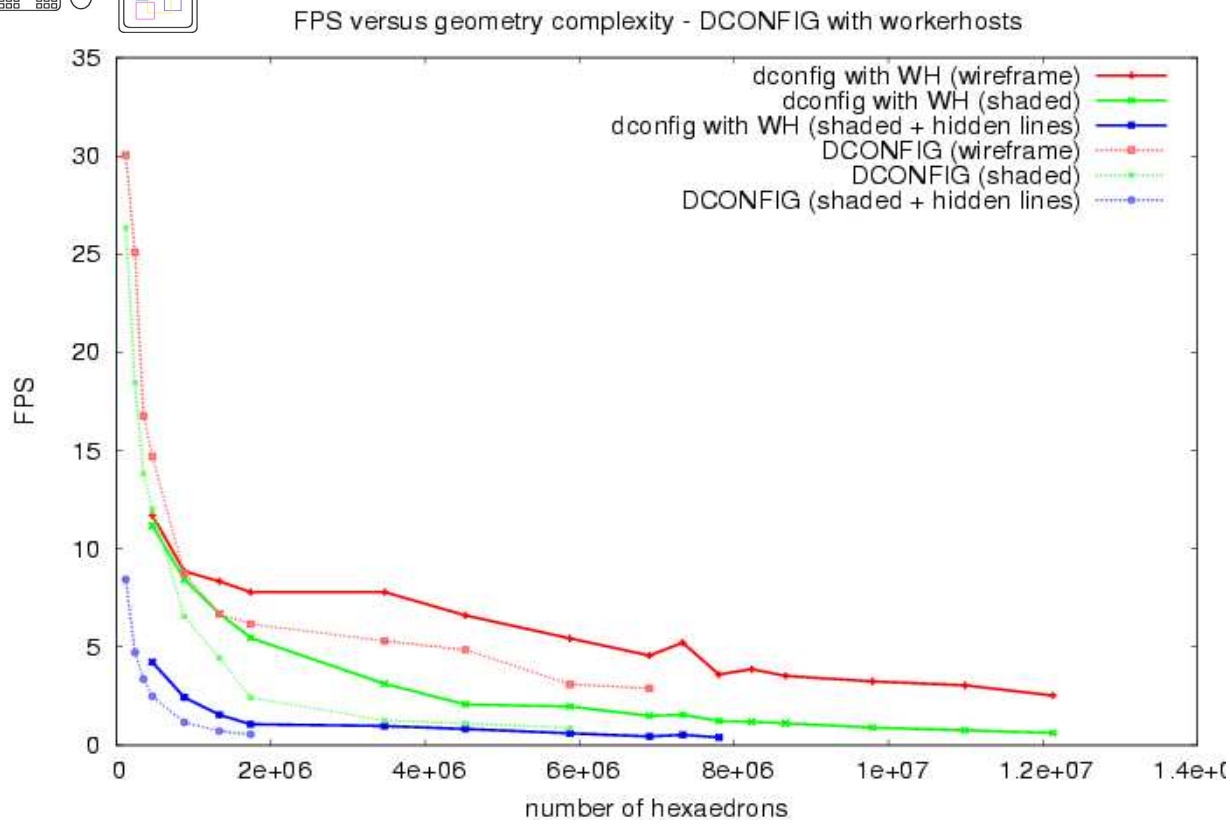


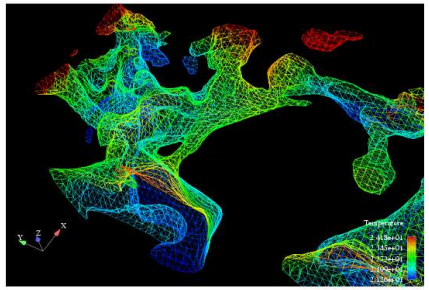


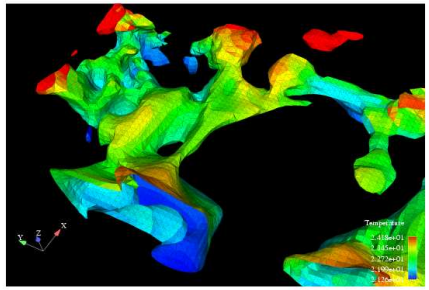


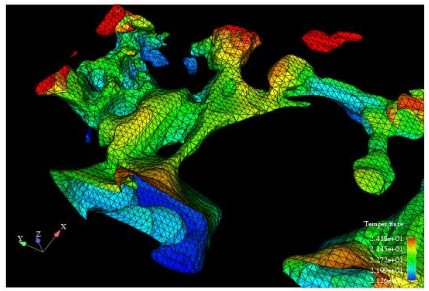
Benchmarks, DCONFIG WH

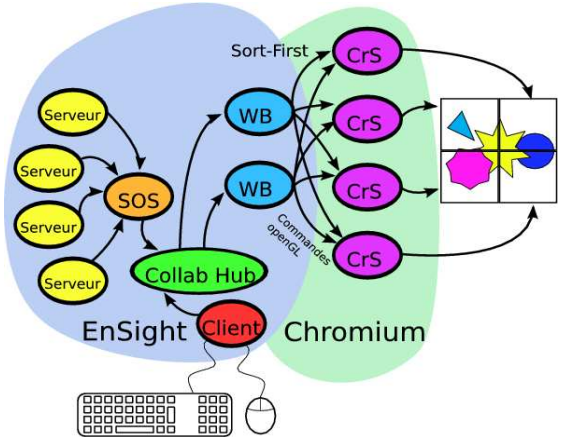




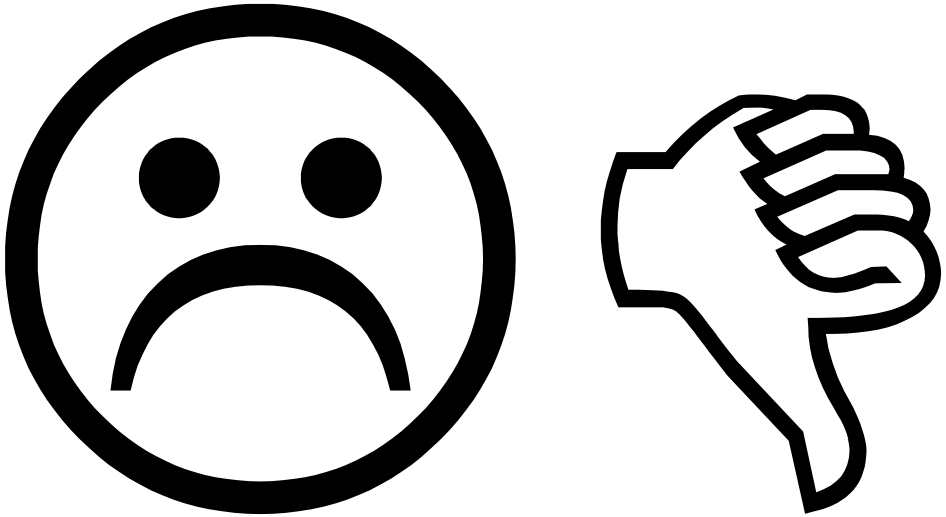


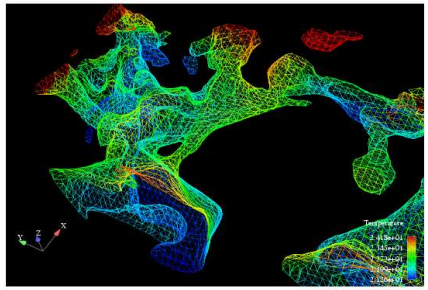


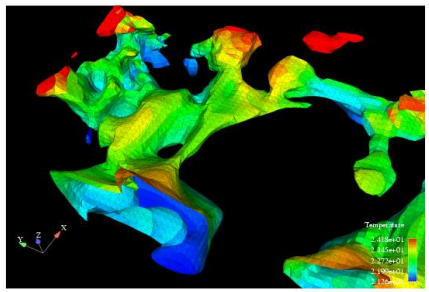


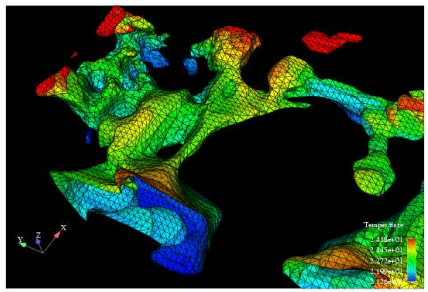


Benchmarks, Cr

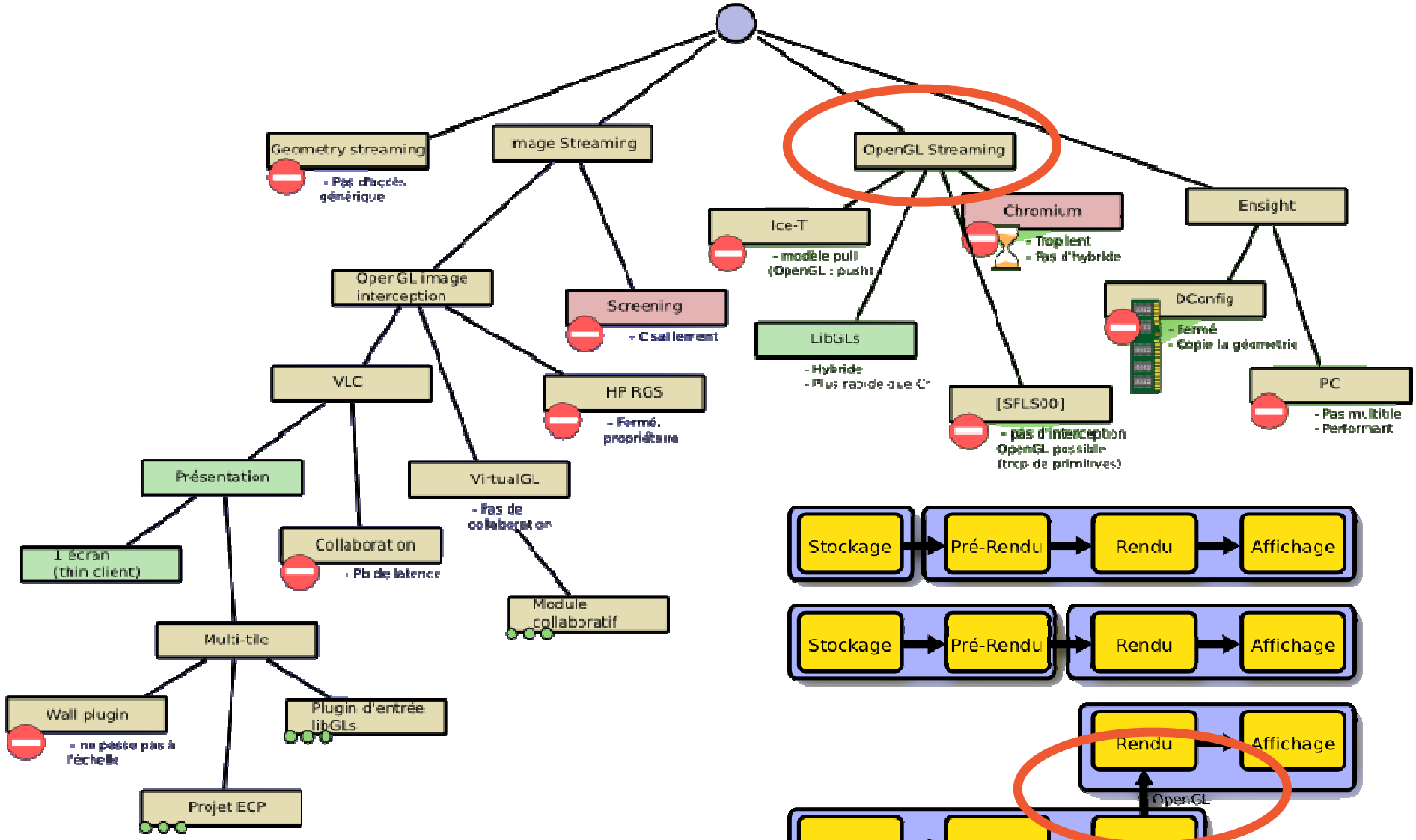








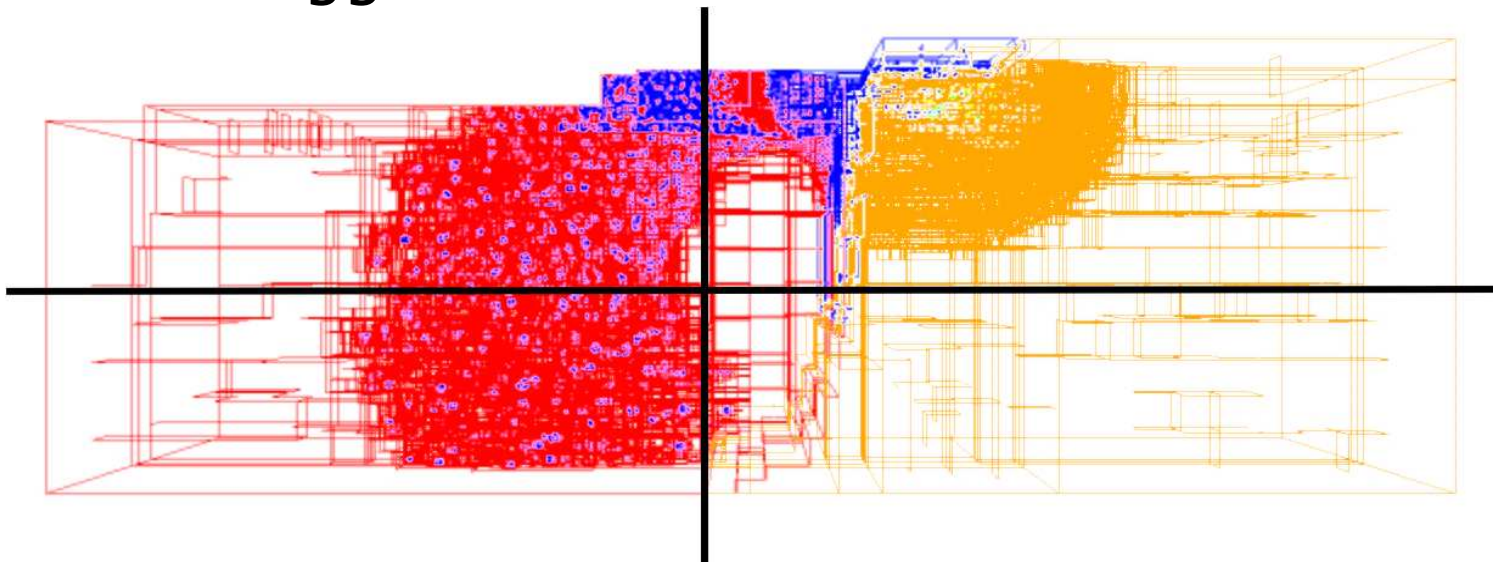
CEA experiments...



libGLs

Ensight DR:

- Closed / proprietary
- Total RAM usage linear w.r.t. number of tiles
  - Meteor: 38.4 Go RAM for 4 tiles (56 domains)
- Chromium: the basic reference, very general, but slow and bugged

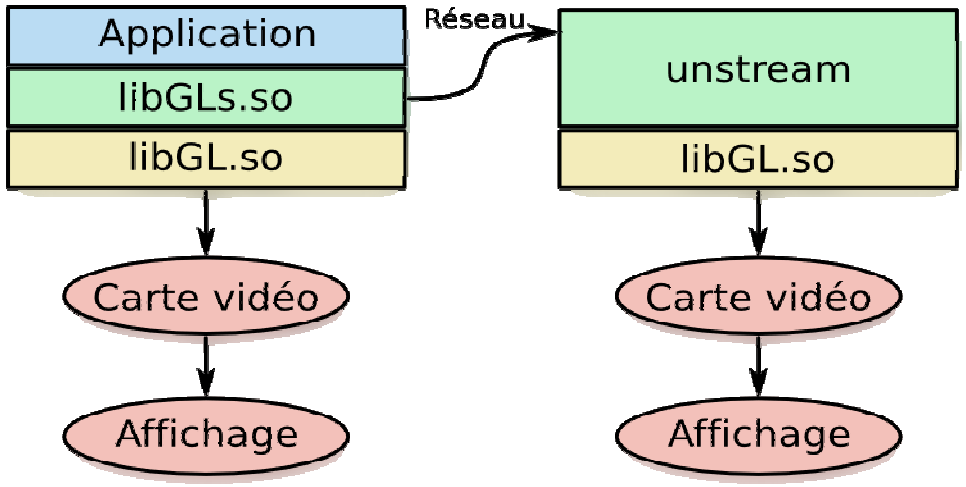


=> libGLs prototype, ligthweight interceptor, hydrid sort-first/last

libGLs

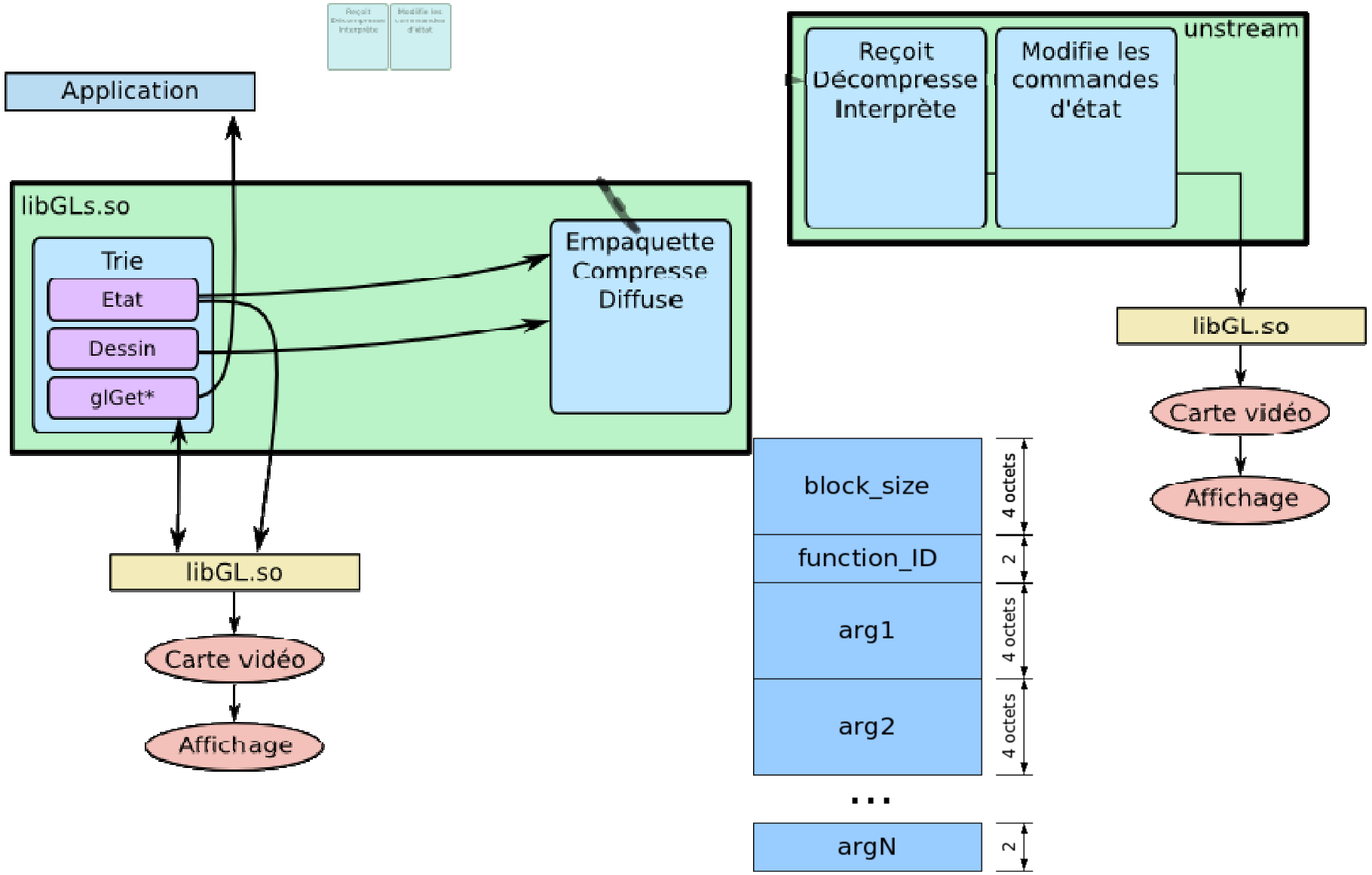
Goals

- Open source: GPL components
- Transparency: compliant with EnSight, but also other viz packages

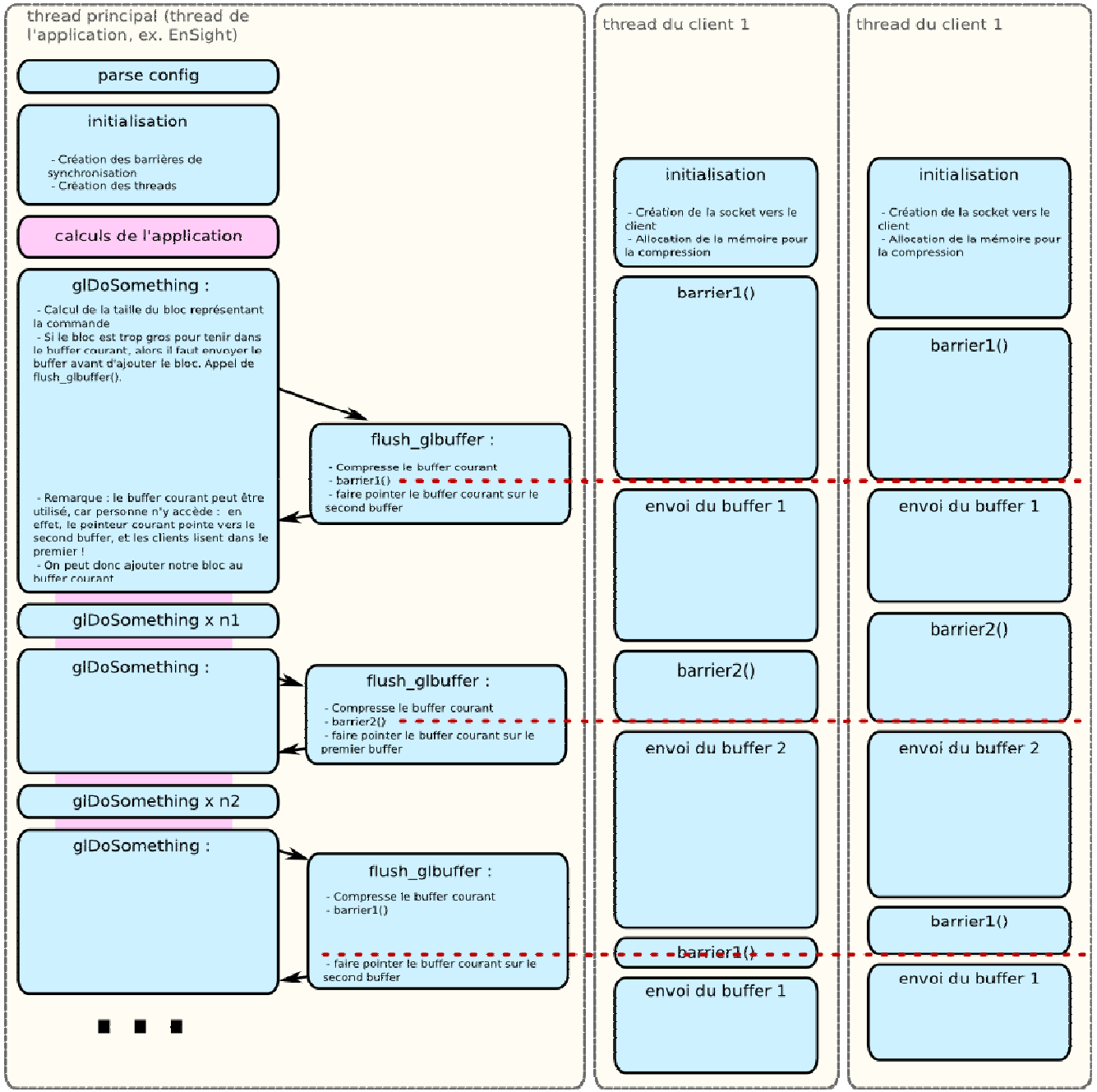


- RAM friendly: streaming ≠ data copy
- Performance: less general than Cr but better suited to our problem

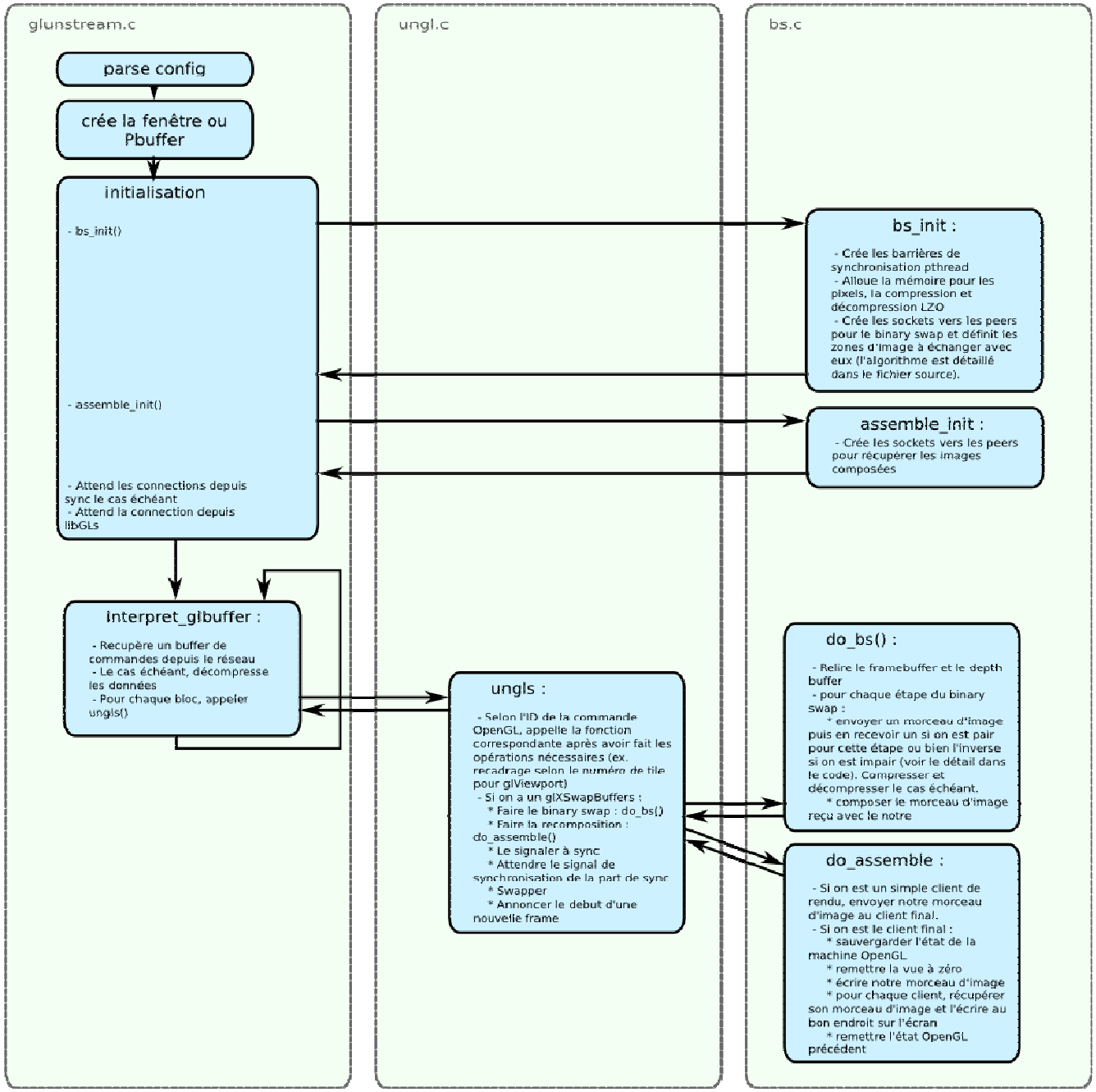
libGLs : command management



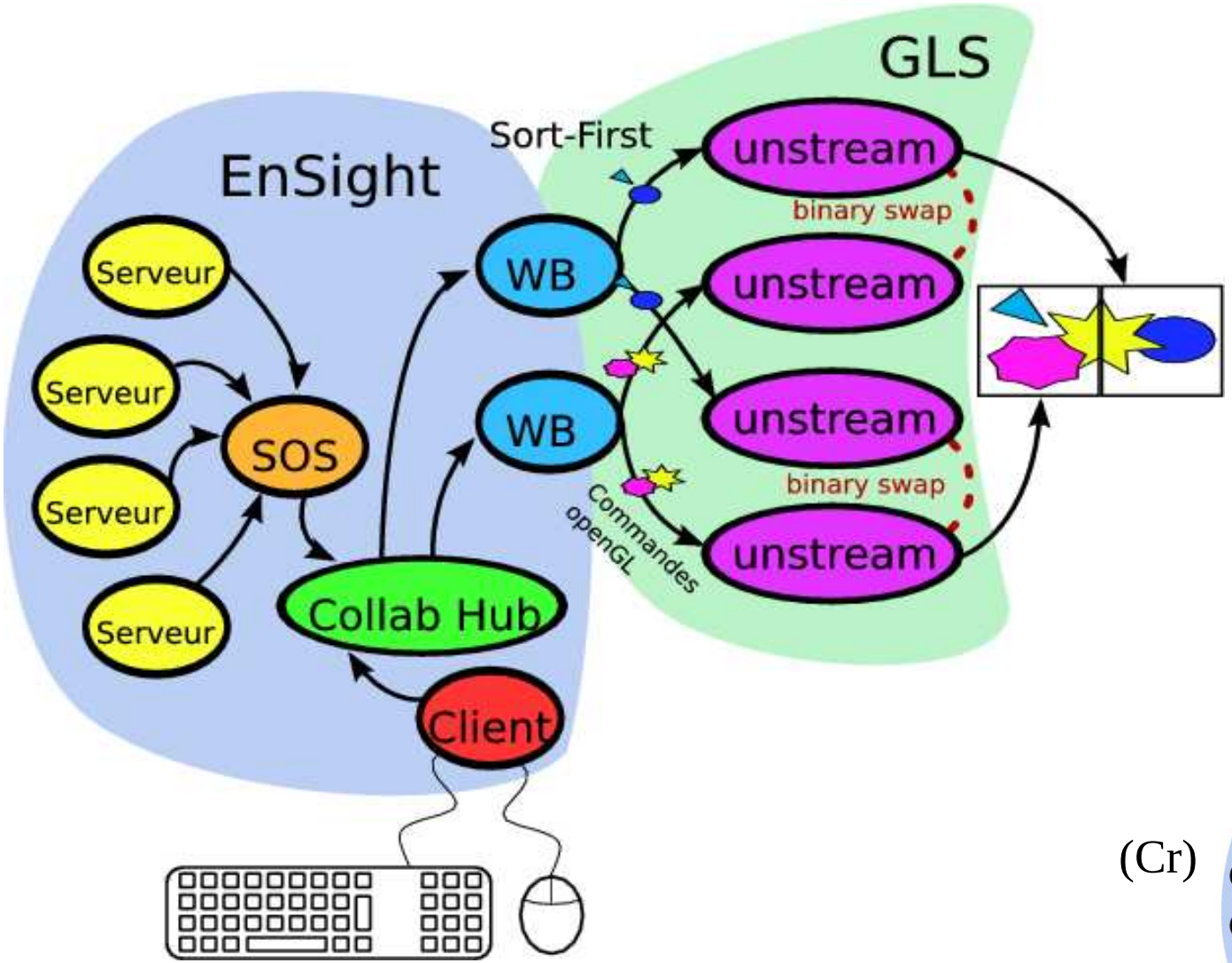
LibGLs (libgls.so)



LibGLs (libgls.so)

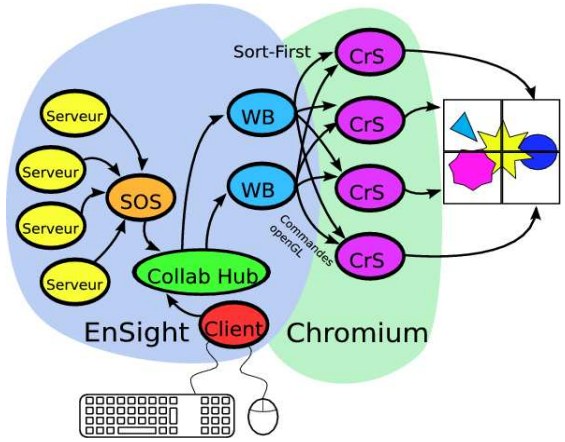


A ligthweight OpenGL interceptor

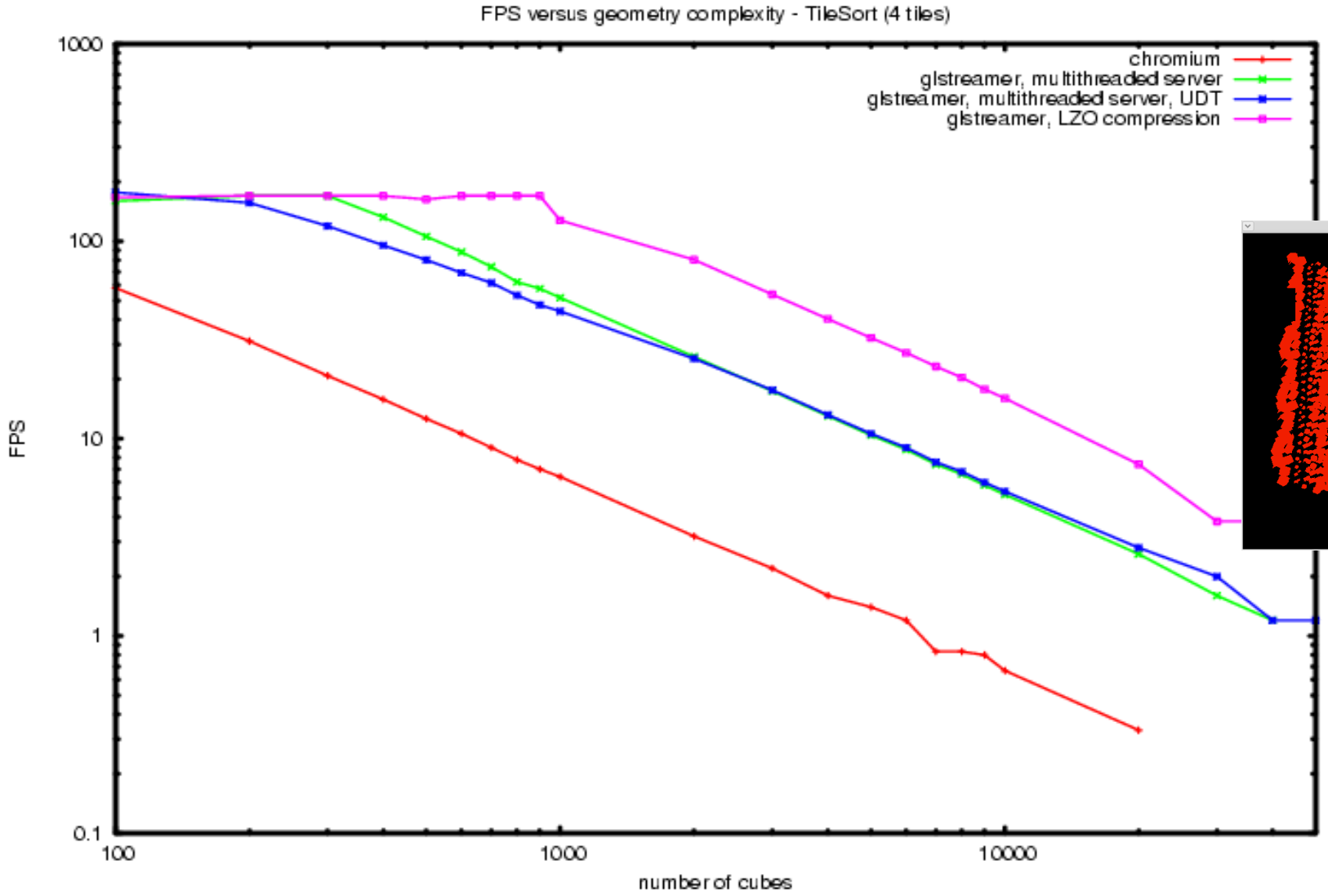


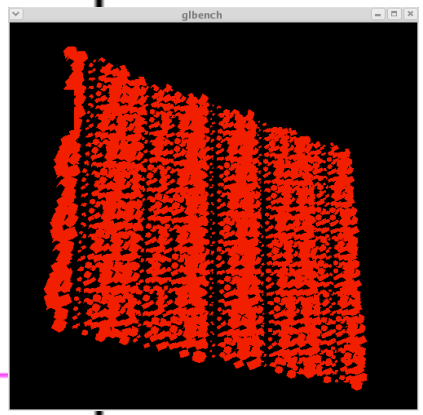
- Géométrie distribuée (par EnSight)
- Sort-First + Last (passe à l'échelle en complexité et en résolution)
- Occupation mémoire totale constante (approche flux)
- Temps de rendu limité par le réseau

(Cr)

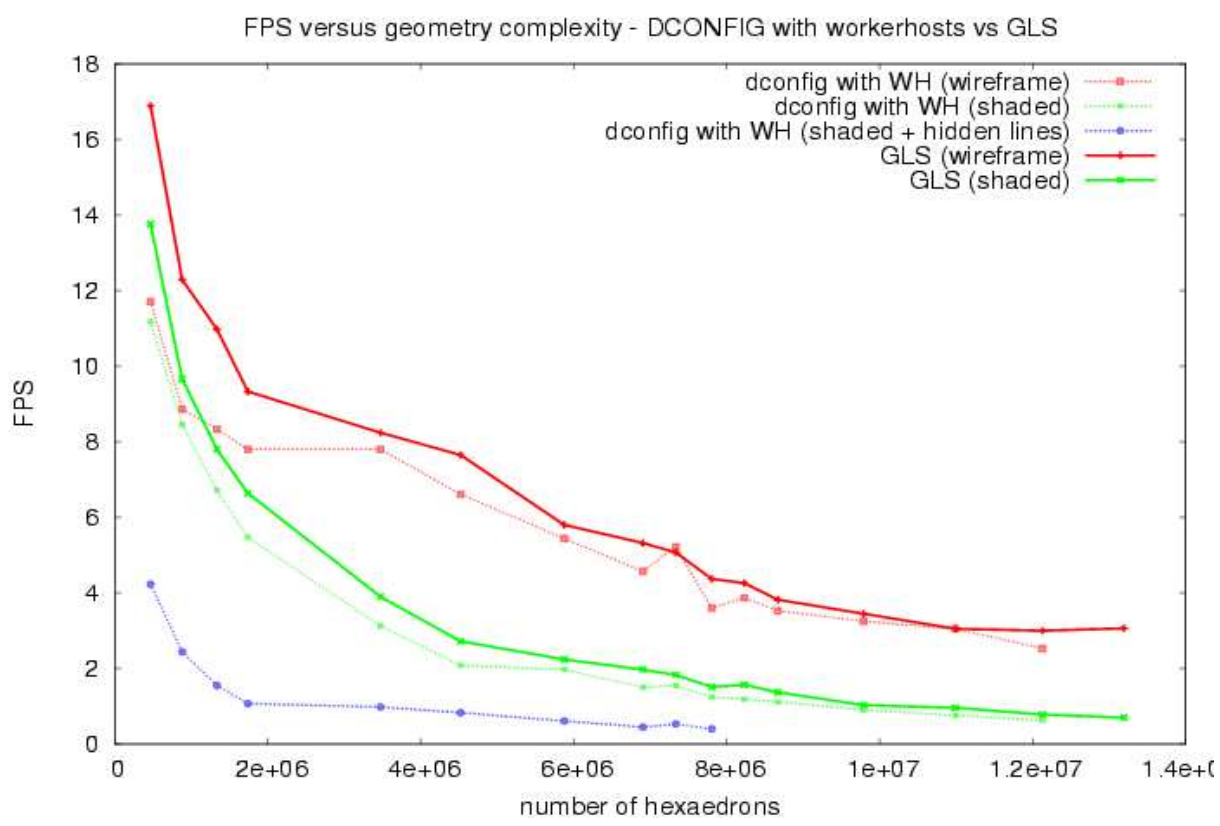


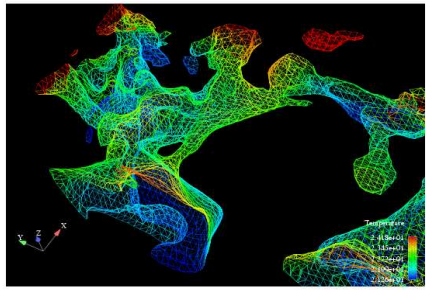
Comparison with Chromium (bench: glbench)

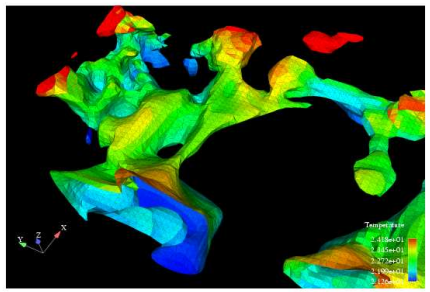


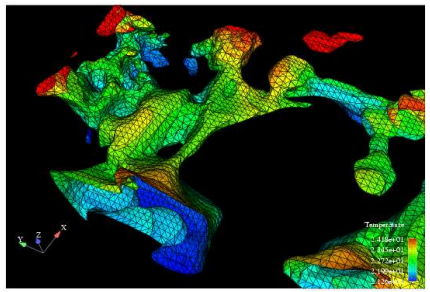


Comparison with DCONFIG 2x4 (EnSight + meteor)





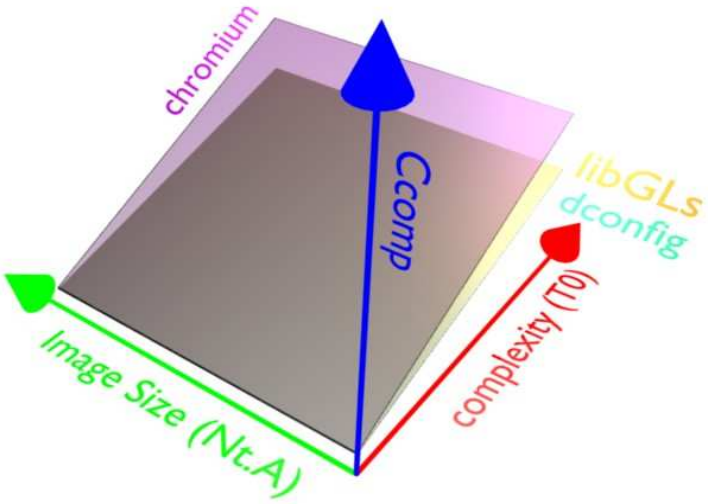


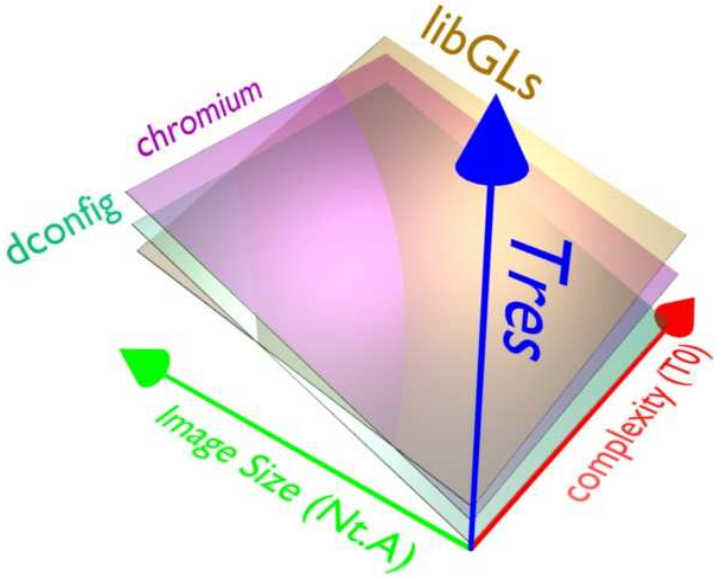


RAM :

- *DCONFIG : 38 Go*
- *LibGLs : 21 Go*

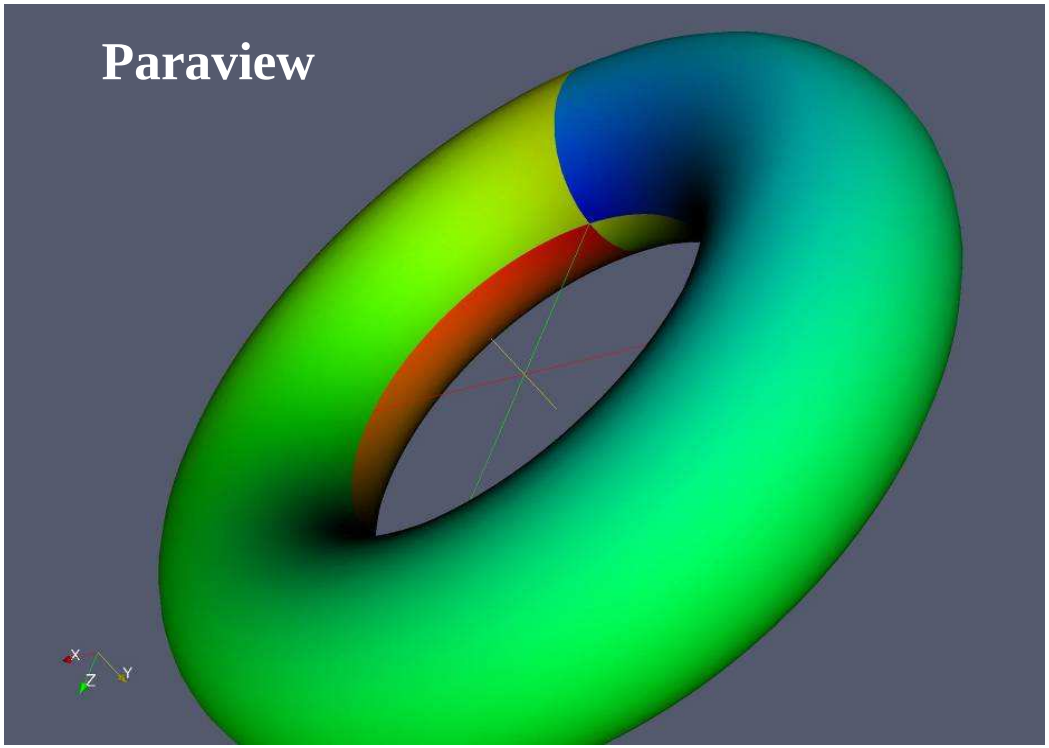
Theoretical comparison with *DCONFIG and Chromium :*

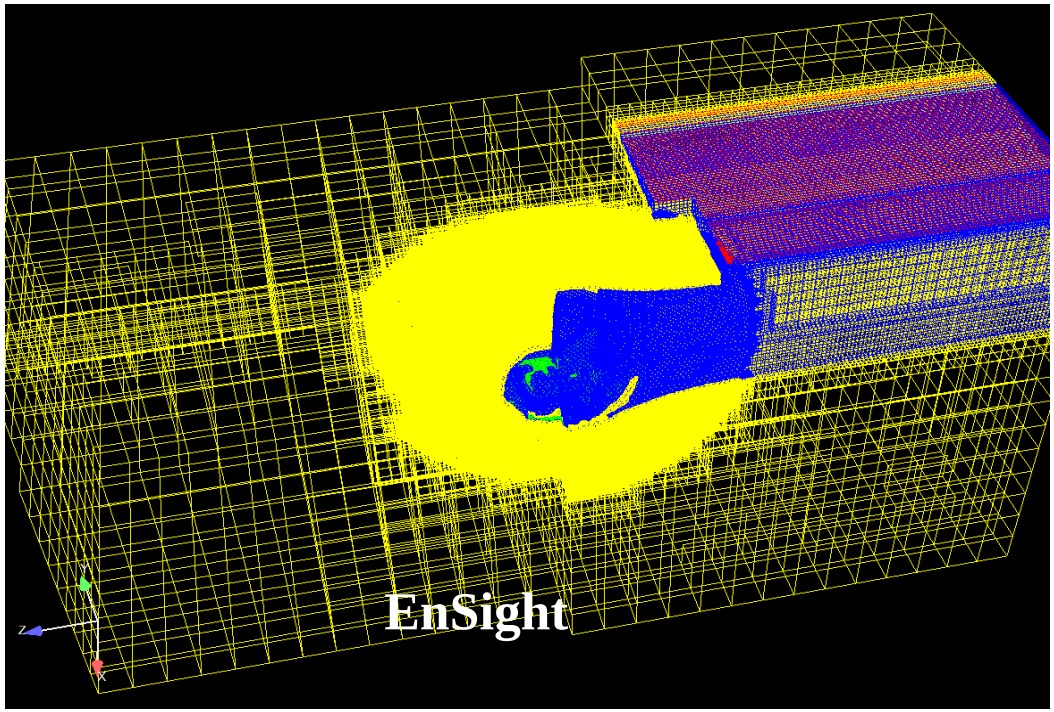




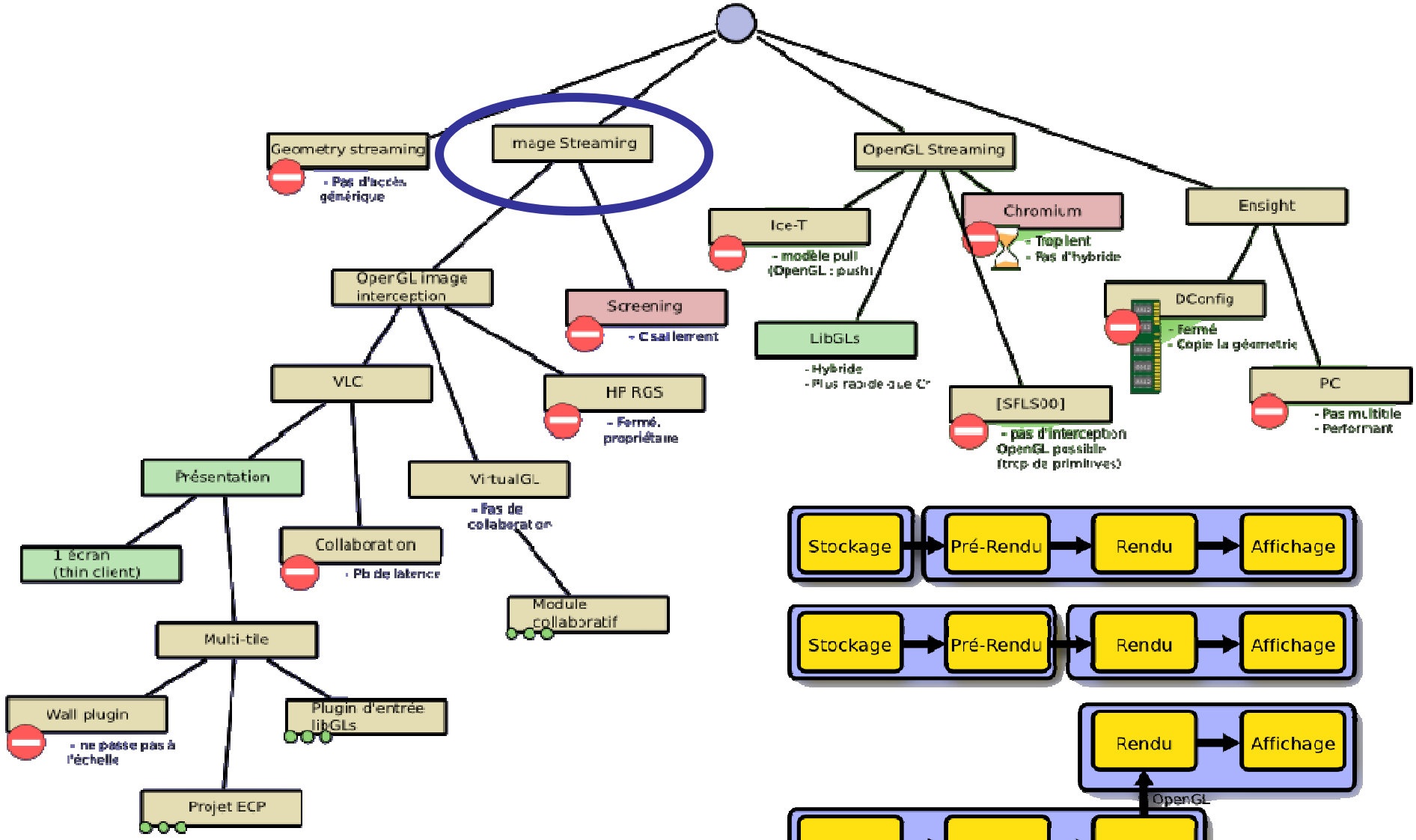
	dconfig	**libGLs**	**Chromium**
Réseau	$O(N_T N_{wh} A)$	$O(T_0 N_T r_c + N_T N_{wh} A r_i)$	$O(T_0 + N_T N_{wh} A)$
Calcul/rendu	$O(\frac{T_0}{N_{wh}} + log_2(N_{wh})A)$	$O(\frac{T_0}{N_{wh}})$ ou $O(\frac{T_0}{N_{wh}} + log_2(N_{wh})A)$	$O(\frac{T_0 N_T}{N_{wh}})$ ou $O(\frac{T_0}{N_{wh} N_T} + log_2(N_{wh})A)$
Mémoire	$O(T_0)$	$O(T_0)$ ou $O(1)$	$O(T_0)$ ou $O(1)$

libGLs : transparent, generic



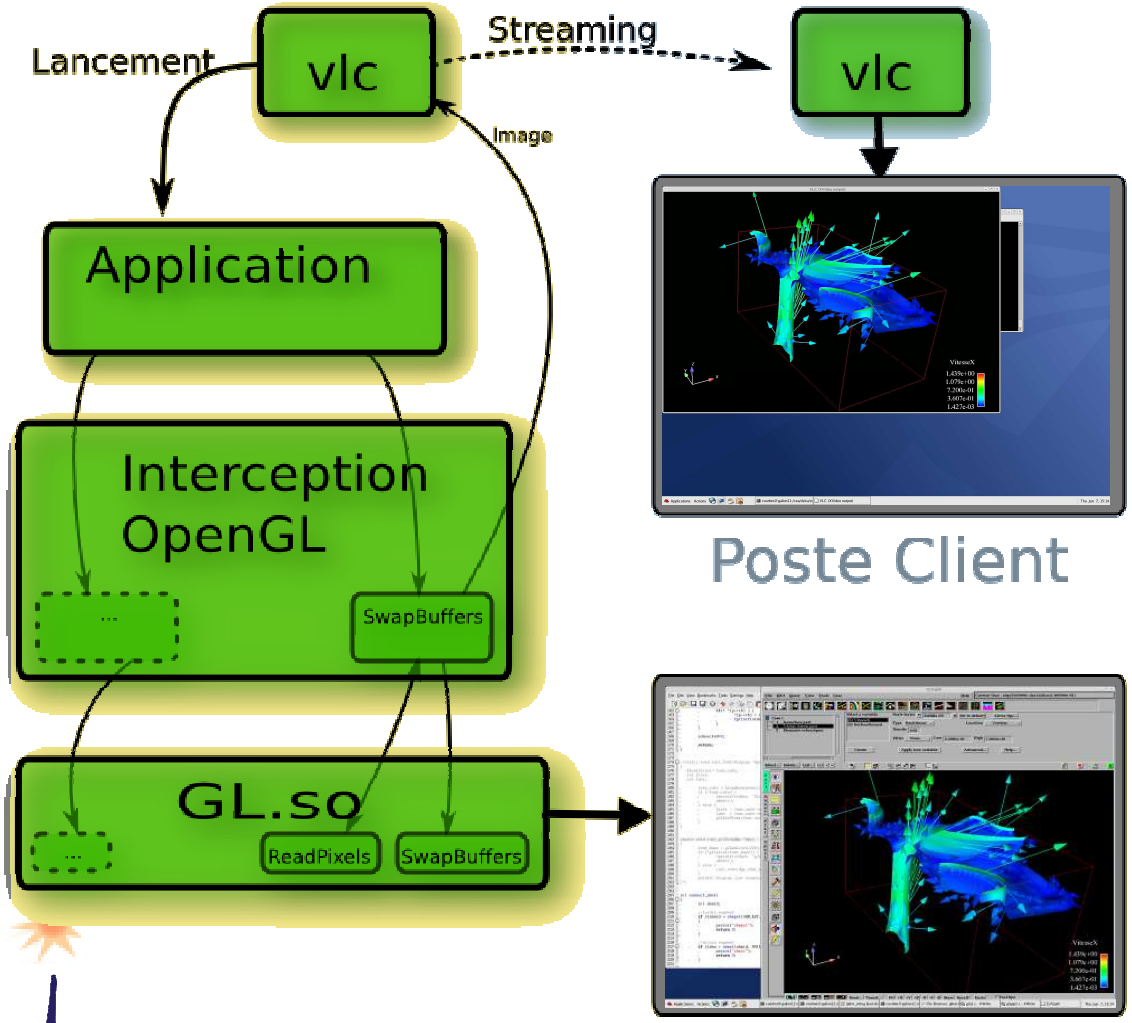


CEA experiments...



Remote visualization (CEA, ECP)

Trying VLC as a remote display engine

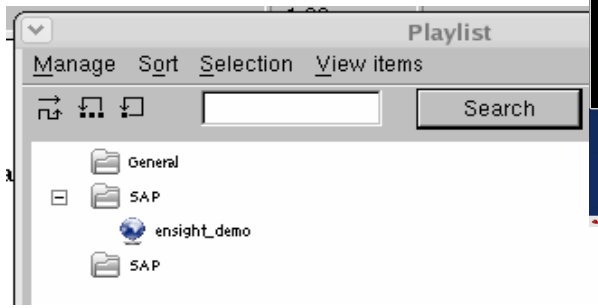


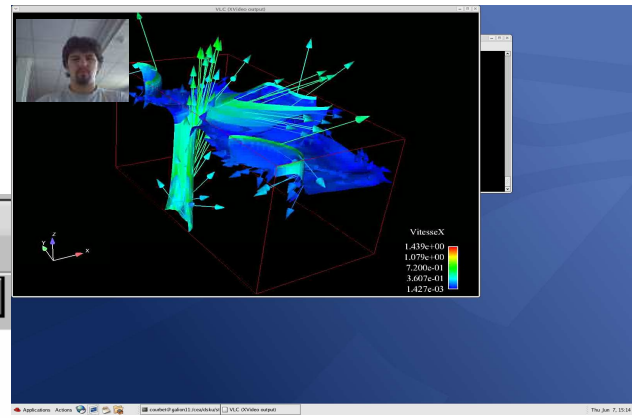
- Server side :
  - Une version modifiée de VLC
  - Interception lib
  - app (ex: EnSight)
  - OpenGL
- Client side :
  - Multimedia reader (e.g. VLC)

VLC steams the framebuffer contents via the network

Results - VLC

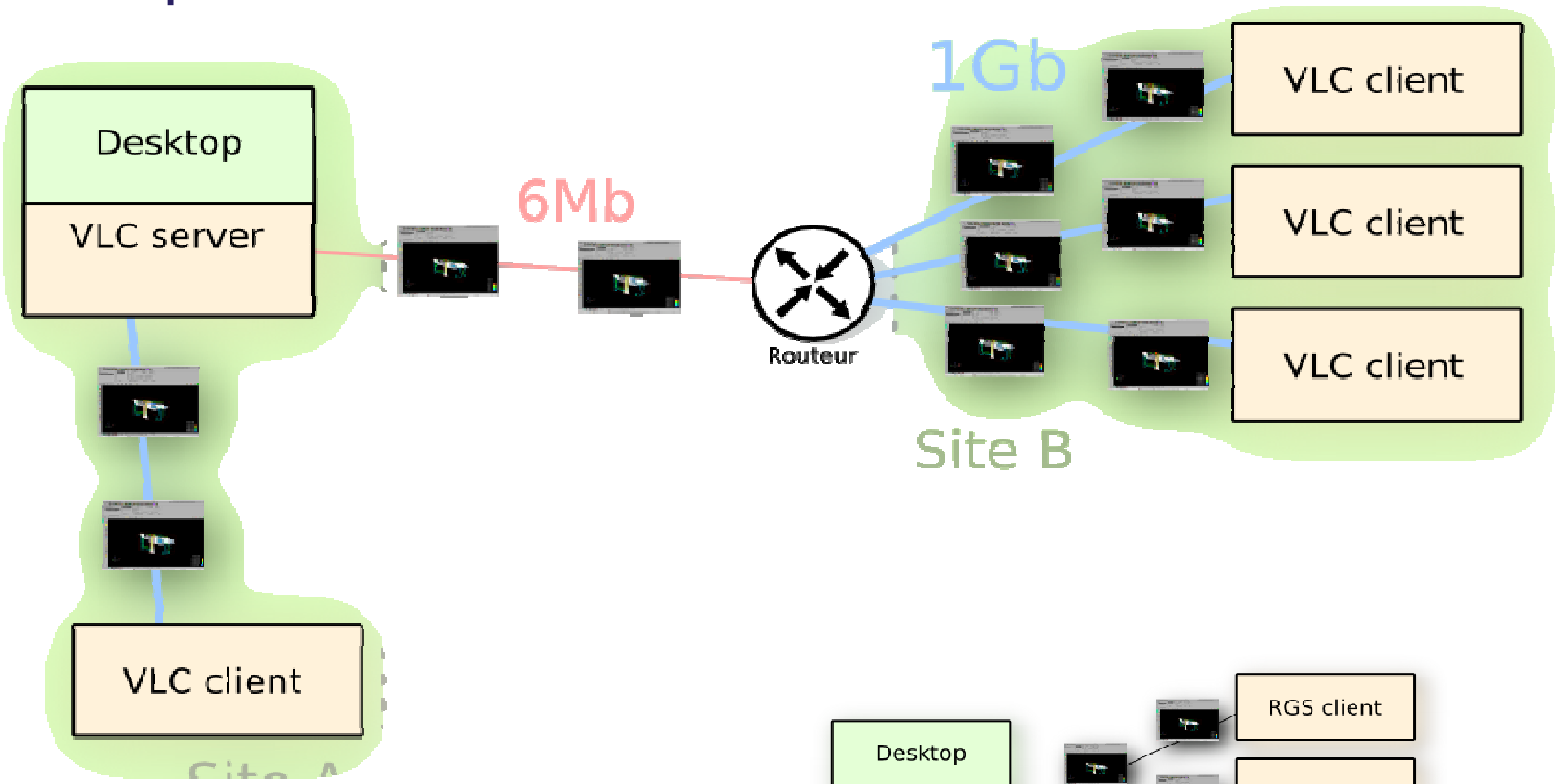
- Advantages
  - Transparent (application unchanged), generic
  - Client side: multimedia reader only
  - Video codec (≠ HP2: image codec): interframe redundancy
  - Video rate and quality tuning (choice of codec and/or bitrate)
  - Other streams can be added:
    - Sound
    - Webcam (visio-conf)
    - Subtitles
  - GPL license
  - Modular
  - Low CPU workload
  - Multicast





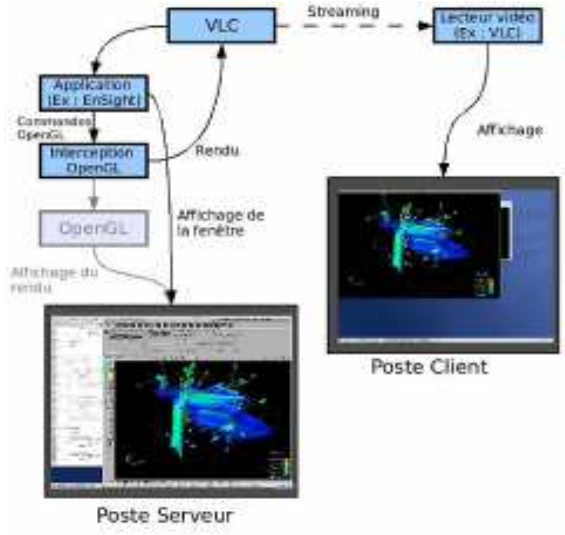
Results - VLC

- Multicast: optimal use of site links



VLC

- Limitations
  - Most of workload on server (application+compression)
    No distribution flexibility
  - MP4V: lower quality than HP2 (RGS)
  - Latency (~1s)
- Collaboration can be handled
  - Add modules on server and client sides
- Web service can be handled
  - VLC plugin for web portal
  - Java applet for collaboration
- ECP is possibly going to further study:
  - the latency issue (profiling in a MT mode…)
  - the parallel streaming possibilities (several rendering streams in case of parallel rendering)
  - other encoders ?



Part 3 : Future work

NextGen VLC

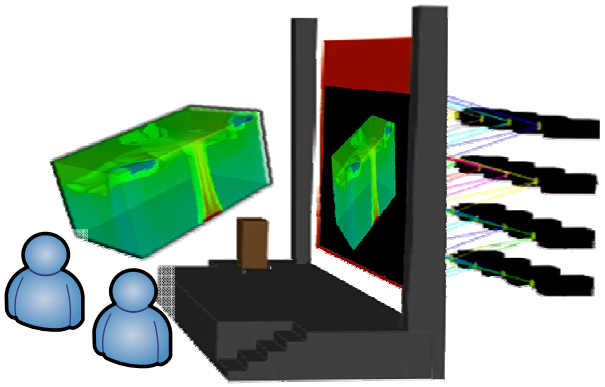
VISUPORTAL HD

2009 : « virtualisation » of the CARRIOCAS R&D technologies

R&D : « NextGen »VLC

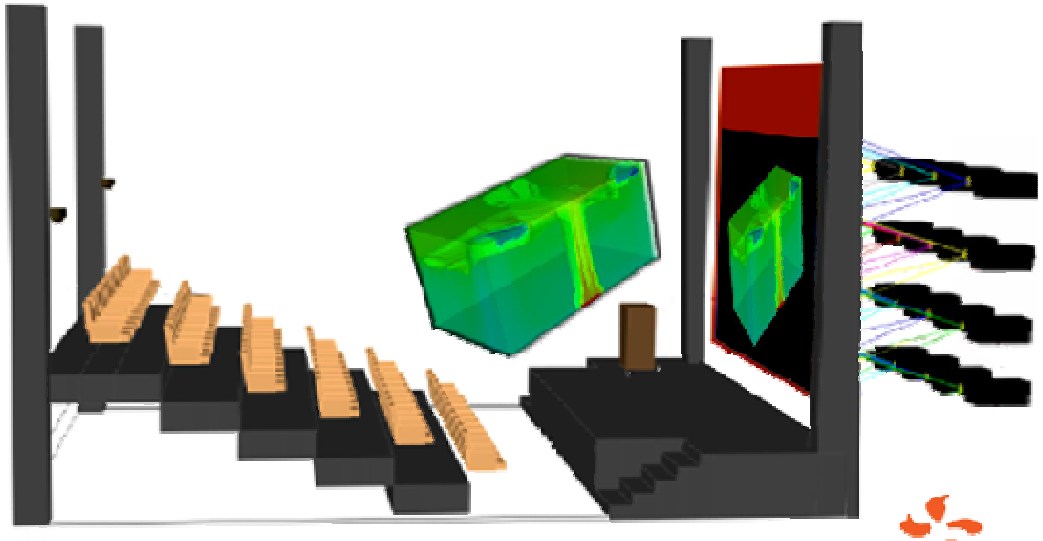
Contributors : ECP with CEA and OXALYA

- New conception/architecture
  - Optimisation of « streaming video»
  - Realtime MPEG 4 encoding
  - Parallel streams ?
- Collaborative mode
- Easy deploiement plugin
- Support for every display :
  - HR display
  - Classic display



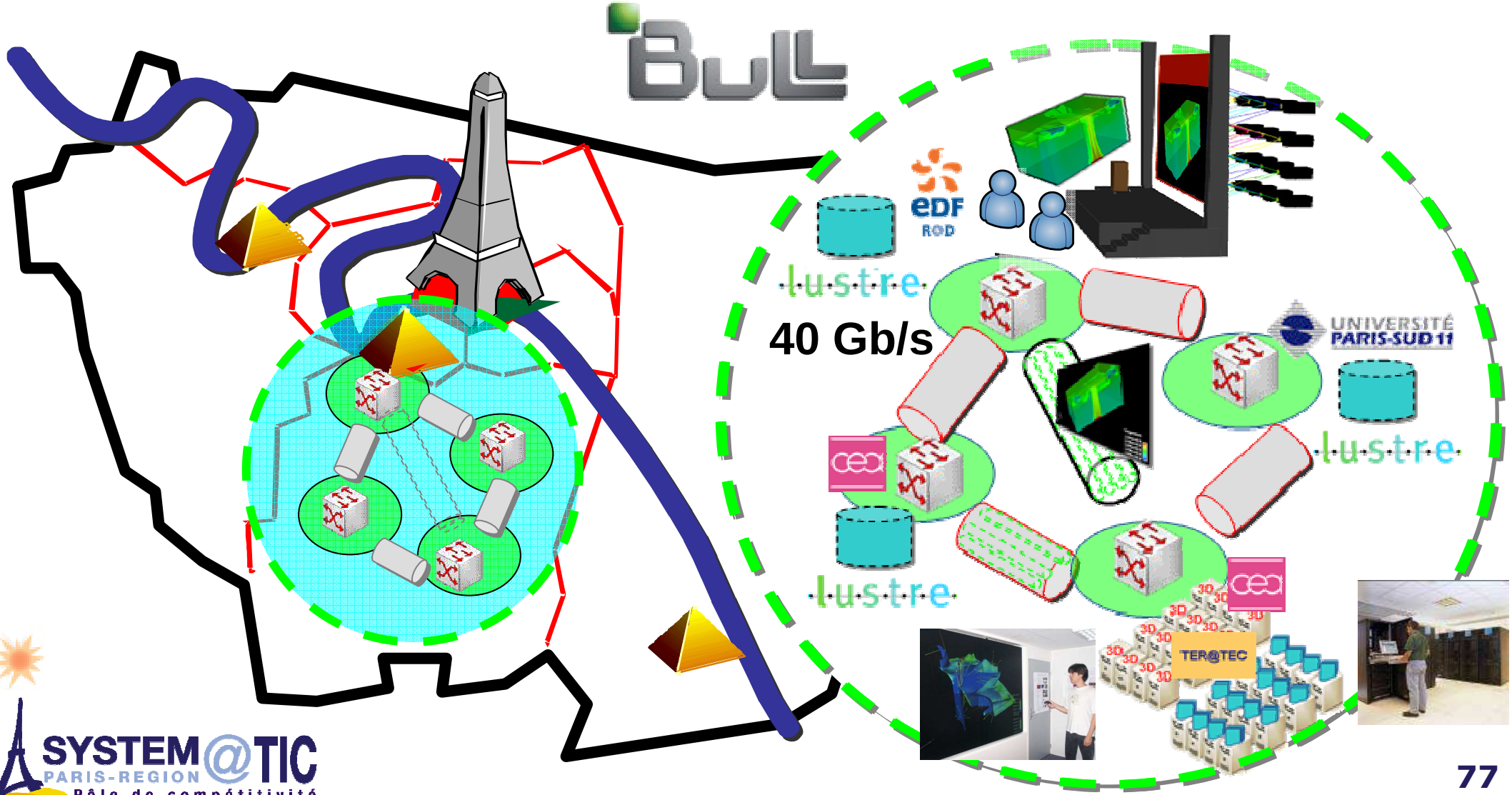
VISUPORTAL HD for High Resolution Display

- R&D « VISUPORTAL HD » at EDF and CEA with OXALYA
  - On Compute & Graphic clusters to manage
    - MIRAGE Jr (CEA-TER@TEC experimental HR display)
    - EDF HR display which will be installed 1st semester 2008 in Clamart (92).
  - Supporting EnSight Gold, DR and other very valuable opensource software packages



Final installation of the remote technical infrastructures for the final demo!

- **2009** : Time to « Virtualize » all the validated and implemented technologies with BULL



Conclusion

- This Visuportal experiment is definitely a tremendous result for the CARRIOCAS team
- **So thank you again to all the colleagues EDF and partners : CEA, ECP, OXALYA!**

- EDF is leading the R&D in the CARRIOCAS project of the future of using high performance visualization ressources. EDF shares this view with other industrials : AIRBUS, EADS, RENAULT and Academics : CEA, INRIA, IFP

- **Don’t hesitate to contact us !**

- **The partners will be pleased to have your comments and questions**

Contacts :
CEA : Dominique RODRIGUES, Jean-Philippe NOMINE
ECP : Christian SAGUEZ, Céline HUDELOT
EDF : Jean-Yves BERTHOU, Christophe MOUTON
OXALYA : Alban SCHMUTZ, Thibaut LAURENT

- Voir aussi www.carriocas.org

SYSTEM@TIC PARIS-REGION Pôle de compétitivité