Héméra
Inria Large Scale Initiative

https://www.grid5000.fr/Hemera

Christian Perez
Avalon

and many co-authors

Motivations

- Scientific issues
  - Large scale, volatile, complex systems
    - Performance, fault tolerance, scalability, data storage, programming models, algorithms, resource management, energy efficiency, etc.
    - Methodological challenges
- Positioning
  - Mathematics,
  - Simulation
  - Emulation
  - ***Experimental testbed (Grid'5000)***
  - Production environment

Outline of the talk

- Overview of Héméra
- Managing Challenging Experiments on Large Scale Systems
  - Some Scientific Challenges of Héméra
- Conclusion

Overview of Hemera

- Goals
  - Demonstrate **ambitious up-scaling** techniques for large scale distributed computing by carrying out **several dimensioning experiments** on the Grid'5000 infrastructure
  - **Animate** the scientific community around Grid'5000
  - **Enlarge** the Grid'5000 community by helping newcomers to make use of Grid'5000
- Open to everyone

Hemera: Participant List

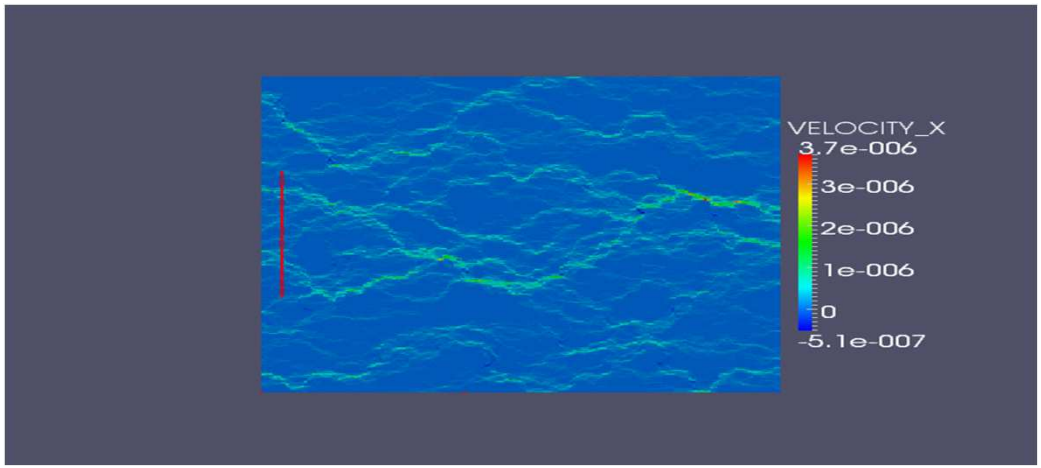
1. ACADIE - Assistance à la Certification d'Applications DIstribuées et Embarquées
2. **ALGORILLE - Algorithms for the Grid**
3. APO - Algorithmes Parallèles et Optimisation
4. ASAP - As Scalable As Possible: foundations of large scale dynamic distributed systems
5. **ASCOLA - Aspect and composition languages**
6. **AVALON - Algorithms and Software Architectures for Service Oriented Platforms**
7. **CC-IN2P3 - Equipe de recherche du Centre de Calcul de l'IN2P3**
8. CEPAGE - Chercher et Essaimer dans les Plates-formes A Grande Echelle
9. **DOLPHIN - Parallel Cooperative Multi-criteria Optimization**
10. GRAND-LARGE - Global parallel and distributed computing
11. ICPS - Scientific Parallel Computing and Imaging
12. **KERDATA - Cloud and Grid Storage for Very Large Distributed Data**
13. OASIS - Active objects, semantics, Internet and security
14. MAESTRO - Models for the performance analysis and the control of networks
15. **MESCAL - Middleware efficiently scalable**
16. **MINC - MIcro et Nanosystèmes pour les Communications sans fils**
17. **MYRIADS - Design and Implementation of Autonomous Distributed Systems**
18. REGAL - Large-Scale Distributed Systems and Applications
19. ROMA - Resource Optimization: Models, Algorithms, and scheduling
20. RUNTIME - Efficient runtime systems for parallel architectures
21. **SAGE - Simulations and Algorithms on Grids for Environment**
22. **SARA - Services and Architectures for Advanced Networks**
23. **SEPIA - Système d'exploitation, systèmes répartis, de l'intergiciel à l'architecture**
24. **ZENITH - Scientific Data Management**

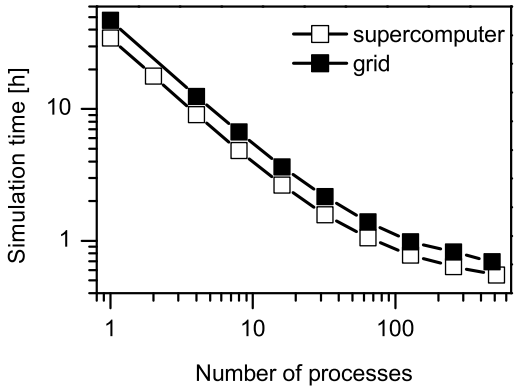
Managing Challenging Experiments on Large Scale Systems

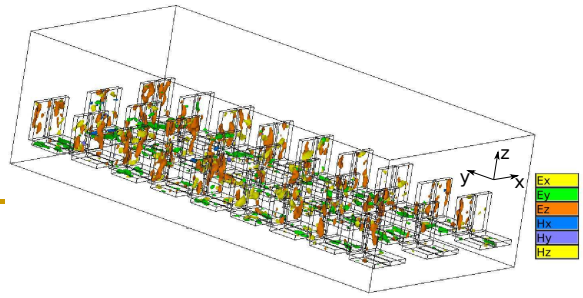
Some Scientific Challenges of Héméra

Supporting Production Usage …

- Application challenges
  - Hydrology (SAGE)
    - Multiparametric 2D stochastic experiments to prepare 3D experiments
    - How to support best effort usage?
  - Electromagnetic Simulation of Oversized Structures (LAAS, MESCAL)
    - Rigorous electromagnetic modeling of complex (multi-scale) propagation channels
    - Compare supercomputers to grid solutions







… To Understand How To Complete Challenging Experiments

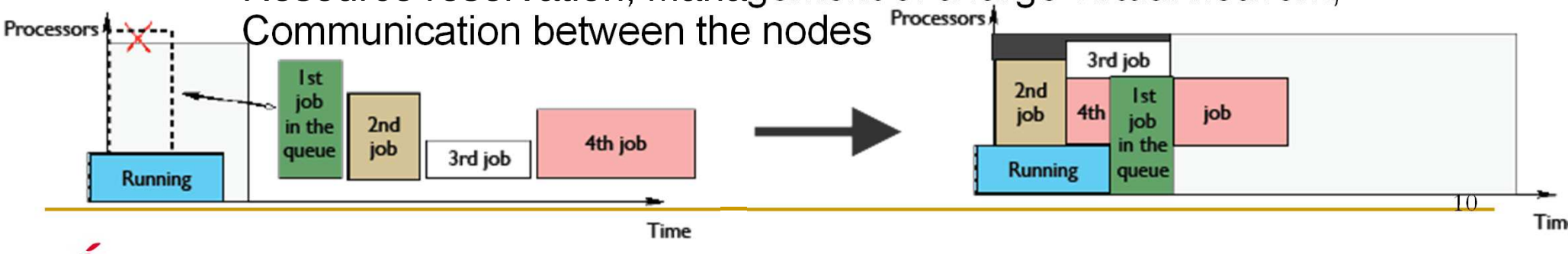
- Towards an Experimental Methodology (Algorille, Mescal)
  - How to describe an experiment?
  - How to check that the platform is well configured?
  - Which data to collect? Experiments? Tools? Platform?
- Axis of work
  - Methodology of the experimentation
    - Scenarios, experimental conditions, metrics, “cahier de laboratoire”
  - Tools for the experimentation
    - Increasing the confidence in experimental results
      - From low level to experiment specific languages (DSLs)
  - Realis : Reproductibilité expérimentale pour l’informatique en parallélisme, architecture et système (ConPas’13)

Deployment Grids and Clouds on Grid'5000

- gLite on Grid'5000 (Algorille, Avalon, CC IN2P3)
  - Designed a set of tools to instantiate a gLite Grid on Grid'5000
  - Enable to validate the behavior of a production tool
  - Ongoing work to further automate this deployment using an experiment orchestration framework
- OpenStack on Grid'5000 (Algorille)
  - Automatize the deployment of OpenStack on Grid'5000
  - Designed a set of tools to instantiate an OpenStack cloud on Grid'5000
  - Already used by an Inria startup (Harmonic Pharma)
    - Evaluate opportunities regarding data processing in the Cloud

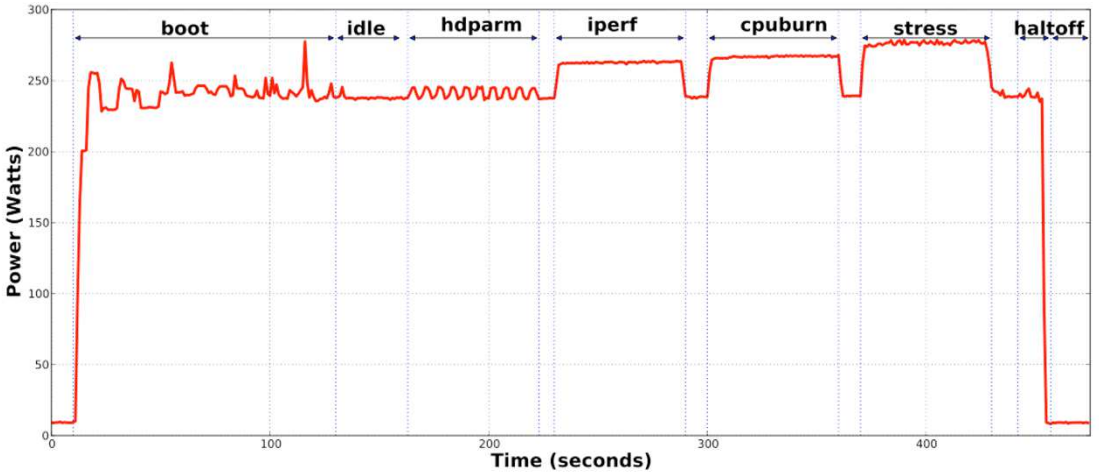
Deploying and Managing a Many Virtual Machines

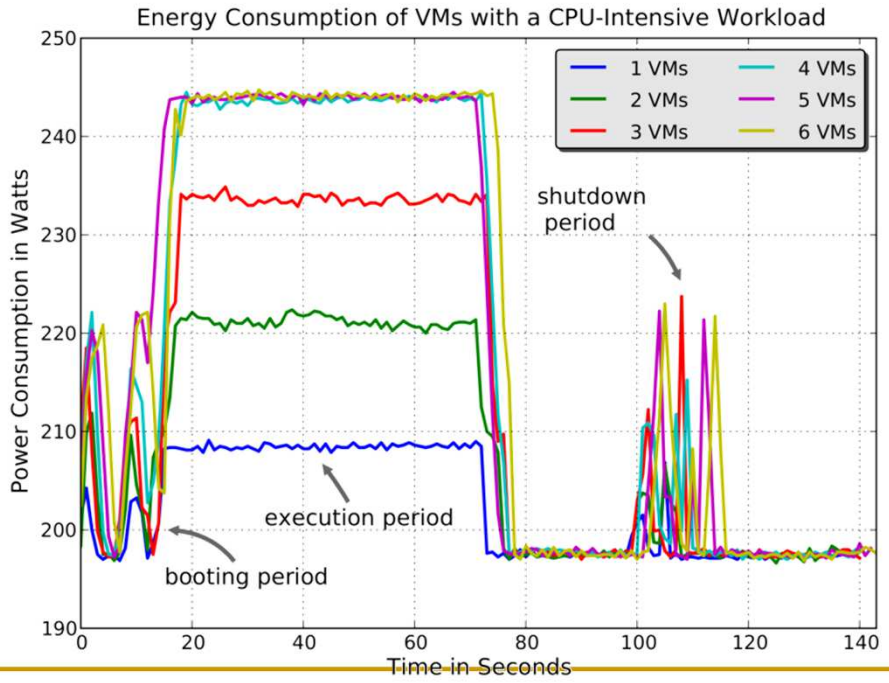
- Deployment (Ascola)
  - Nation-wide management of virtual machines over Grid'5000
    - 5 sites on 11 clusters (KaVLAN)
- Management (Ascola)
  - Management of **10000+** Virtual Machines on 512 physical machines
    - VM can be moved on another site using live migration capabilities
  - Use of advanced mechanisms to satisfy scheduling criterion (load balancing, consolidation, ...)
  - Flauncher: portable tools for deploying many VMs
    - Resource reservation, Management of a large virtual network, Communication between the nodes



Monitoring Energy Consumption

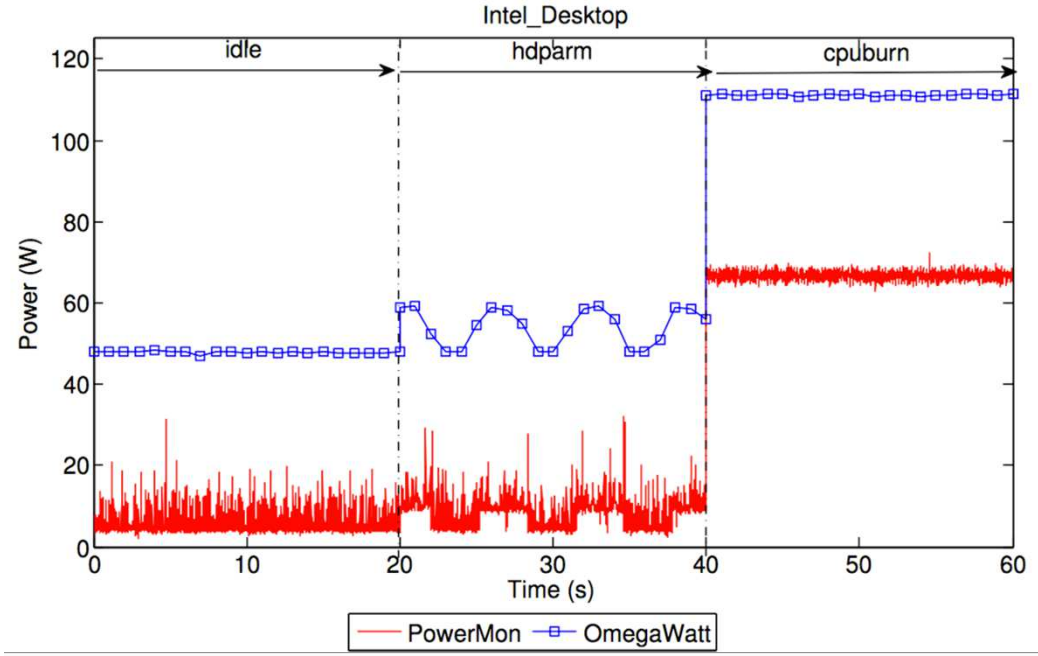
- Develop an eco-system
  - Power sensors
  - Gather information
  - Publish it
- From physical machines to virtual machines
- (Avalon, Ascola, Myriads, IRIT)

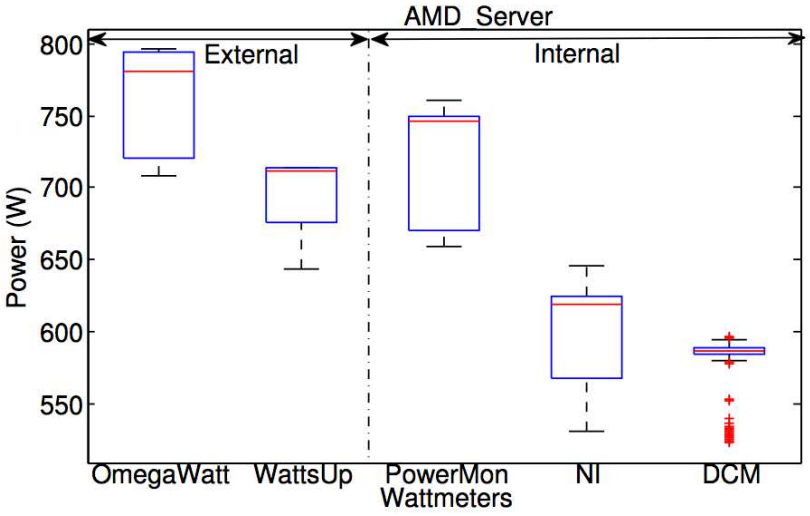


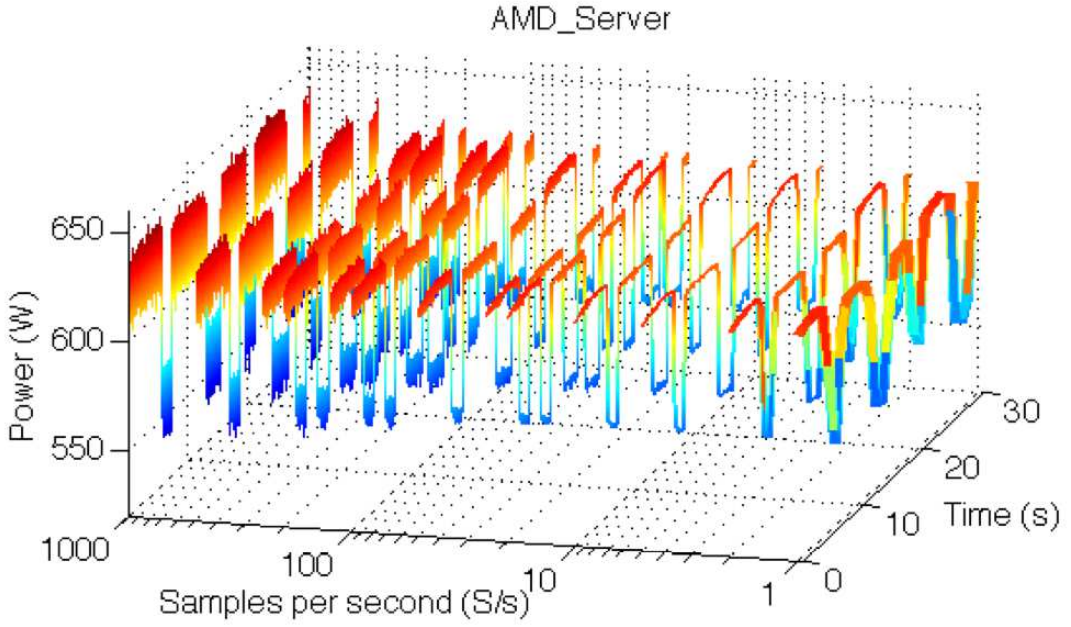


Understanding Energy Measurements

- Measurements
  - Validity
  - Precision
  - Frequency

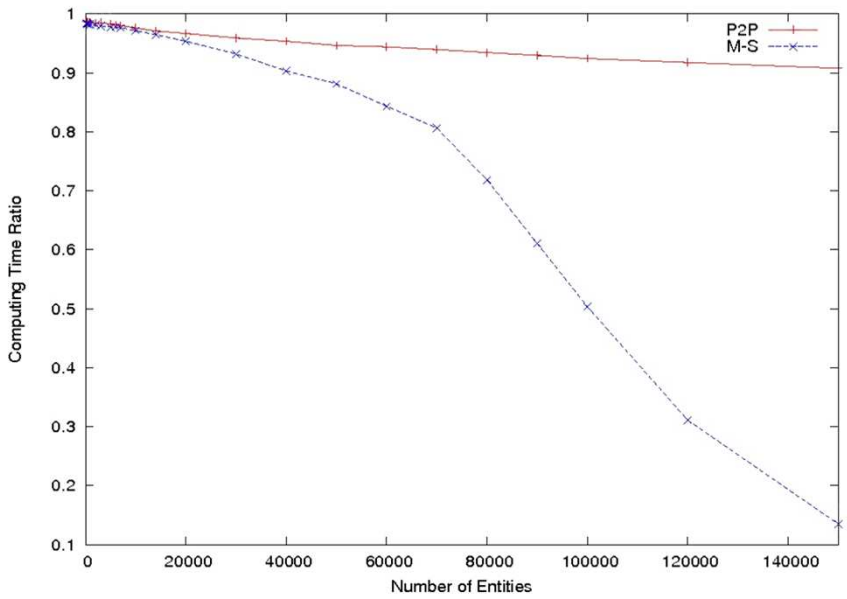






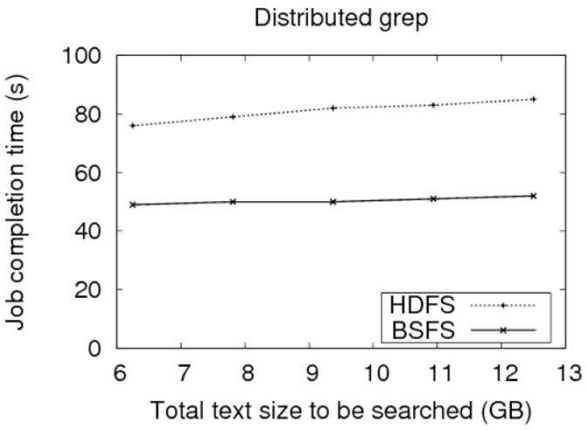
Large Scale Branch & Bound (Dolphin)

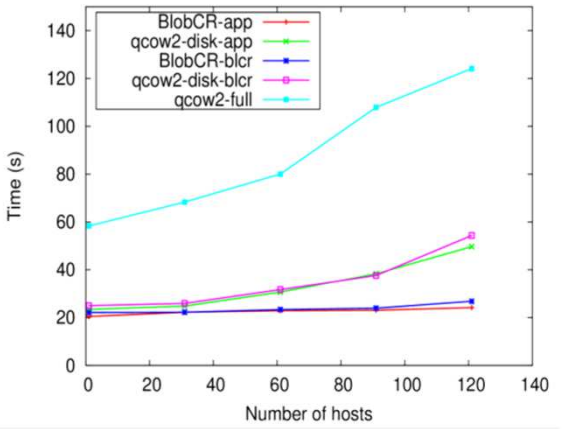
- A new fully distributed B&B
  - Validated using a Pastry-like overlay and up to 150.000 processes
- Towards a Fault-tolerant Peer-to-peer B&B
  - A hybrid two level approach
    - Distributed work sharing and overlay maintenance
    - Centralized Checkpointing
  - Validated under several fault models
- Towards Adaptive Scalable Dynamic LB

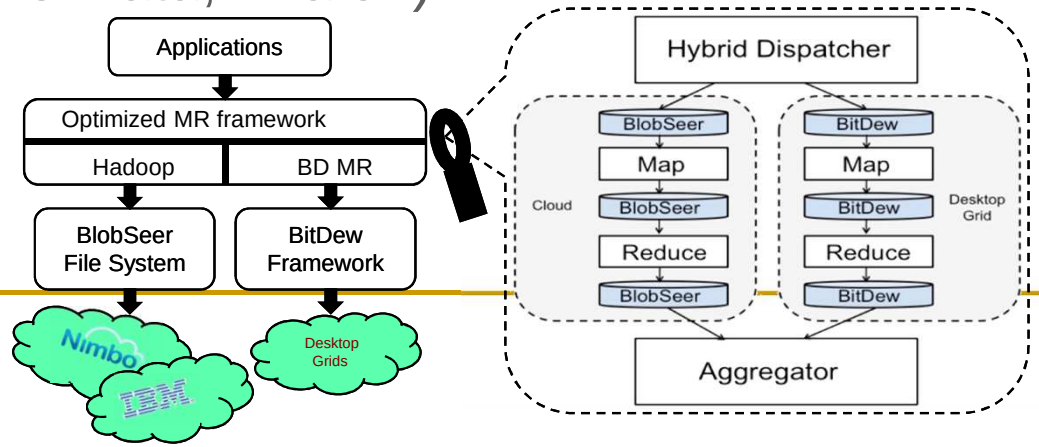


Scalable Distributed Processing Using the MapReduce Paradigm

- Data Management
  - Hadoop performance improved (up to 38%) with BlobSeer File System (KerData)
- VM Management
  - Fault tolerance version-based VM Check-point/Restart (KerData, UIUC)
- MapReduce
  - Optimizing data mining primitives (Zenith)
  - Hybrid MapReduce (KerData, Avalon)







Gather Knowledge About Challenging Experiments on Large Scale

- Experimental methodology
  - Experiments conception, data mining, simulation
- Tools for experimenting
  - System tools
    - Grid'5000 tools such as Kadeploy, kavlan, OAR, etc.
    - Platform description, conformance verification
  - User tools
    - Experiment management
    - Large scale (grid, cloud, etc.) deployment
    - Energy monitoring
    - Data management
- Use Cases – studying complex situations
  - Large Scale Cloud Platform (> 10k VM) + MapReduce + Data Management + Energy Consumption Monitoring

Animation

- Grid'5000 School
  - 2006, 2009, 2010, 2011, 2012 (100, 72, 80, 67, 69 participants)
  - Tutorials, Invited talks, research "papers" (~utilization of G5K)
  - Large Scale Deployment Challenge
- GreenDays events
  - 2011 (Paris), 2012 (Lyon), 2013 (Luxembourg), 2014 (Lille)
- Serie of virtualization workshops
- Serie of workshops around MapReduce/Big Data
  - MapReduce@HPDC (2011, 2012), ScienceClouds@HPDC (2012,2013), VTDC@HPDC (2012, 2013), BDMC @Euro-Par 2012

Conclusion

- **Experimental platforms (and observation instruments) are essential** in the CS methodology - like in other sciences!
- Many research kinds are using Grid'5000
  - HPC, Grids (Classical/Desktop), Clouds, Distributed, Green, etc
    - A validation tool for applications/middleware before going to production
- Hemera
  - Target to solve challenges & to structure the French community
    - 24 teams: 13 core teams (not all Inria), 11 "side" teams
  - Focus on core methods and tools
    - Experimental methodology
    - Tools for experimenting
      - System tools, User tools
    - Use Cases – studying complex situations