Interfaces between Production and Research Grids

Frédéric Suter

Conseil Scientifique de l'Institut des Grilles
1er Juin 2010

Outline

- Research Grids vs. Production Grids
  - Objectives
  - Types of jobs
  - Job Life Cycle
  - Load
  - Resource Sharing
  - Behavior
  - Administration Constraints
- Research Grids and Production Grids
  - The Interface Program
  - Possible Collaboration Points

Different Objectives

Production Grids

- Goal: Feasibility
- Means: Toolboxes
- Keyword: Transparency
- Example: EGEE/EGI
- Execute the applications of today

Research on Grids

- Goal: Performance
- Means: Algorithms
- Keyword: Control
- Example: Grid'5000
- Prepare the environments of tomorrow

Different Types of Jobs

Production Grids

- Driven by LHC and High Energy Physic
- A vast majority of 1-CPU jobs
- 100% on the AuverGrid trace[1]
- Longer jobs ($>$ 7h on average)

Grid'5000

- Driven by the HPC community
- Many parallel jobs
- 53/46 on the Grid'5000 trace[2]
- Shorter jobs ($<$ 1h on average)

[1] `http://gwa.ewi.tudelft.nl/pmwiki/pmwiki.php?n=Workloads.Gwa-t-4`

[2] `http://gwa.ewi.tudelft.nl/pmwiki/pmwiki.php?n=Workloads.Gwa-t-2`

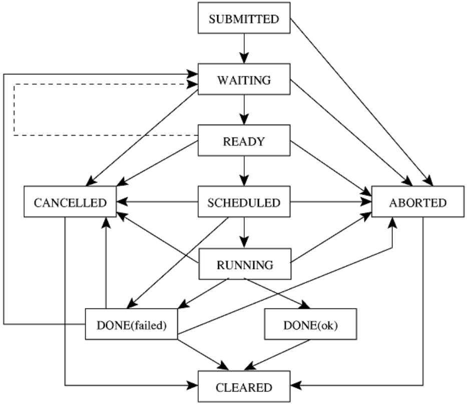
Different Job Life Cycles

Grid'5000 for a 1-proc job

- Look at Monika or the Gantt chart
- Select one site that has at least one node available
- Go there (maybe with my data)
- call `oarsub` (maybe with `-I`)
- get my machine quite immediatly
- Deploy my environment
- run my application
- Get the results or fail

Different Job Life Cycles

On Production Grids (EGEE)



- Before submission, have to specify the requirements (in JSDL for instance)
- No deployment, have to find a suitable node
- The `waiting` state can be neglected

Different Loads

Grid'5000

- Not heavily loaded
  - by design, dimensionning experiments can still run
- Always some nodes available
- Jobs can wait (for hours) if
  - They require many cores
  - They asked for a specific (and demanded) resource
  - There is a big conference deadline

Production Grids

- Always overloaded!
- Example of the IN2P3 Computing Center
  - 10 jobs for 8 CPUs
  - CPUs at full speed 80% of the time
    - Inactivity only due to data staging
  - 8,000 jobs running and 16,000 jobs waiting (for days) in queue

Different Resource Sharing

Grid'5000

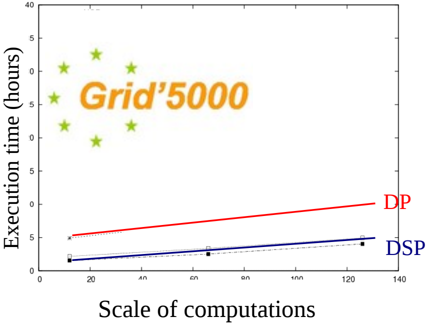
- Once you get an account and have signed the charter
- Do what you want (with respect to the charter)
- Your behavior is traced by Kaspied

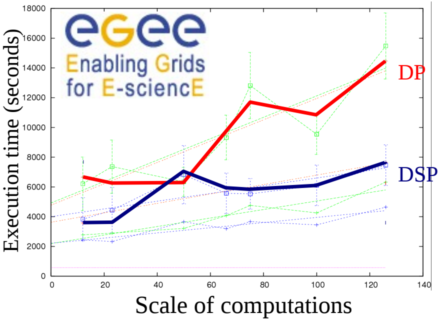
Production Grids

- Rely on VOs (Virtual Organizations)
- You can only access to resources where your VO is allowed
- Sharing among VOs is decided beforehand
- At a computing center level
  - Applications make resource requests each year
  - Consensus has to be found

Different behaviors

- **Performance comparison across platform**
  - Same (workflow-based) application running on reserved resources (G5K) and in production (EGEE)
  - Two parallelization modes (DP and DSP)





Different Administration Constraints

Grid'5000

- Users can deploy their own image
    - Admins "only" have to maintain the default image
- Three critical services: OAR, Kadeploy, and Kvlan

Production Grid

- No virtualization (yet)
- Admins have to maintain
    - Operating System(s)
    - Libraries
    - Middleware stack
    - Licensed software
- Upgrade is a long process
    - 1 year to move from SL4 to SL5 at CC IN2P3
    - 2 or 3 concurrent versions of the OS
    - Scientifics often keep the sources, compiler and even binary
        - To ensure data processing under similar conditions somtimes years after

Conclusion

Grid'5000 is not a (production) Grid

- It's an scientific instrument
- Going to a production mode is not trivial
- Because the focus are different
  - Grid'5000: Controlled environment, you know everything
  - Production Grid: execution platform, (almost) everything is hidden

Outline

- Research Grids vs. Production Grids

- Research Grids and Production Grids

  The Interface Program

  Possible Collaboration Points

Interfaces

Making connections

- One of the missions of the Institut des Grilles
- In cooperation with Aladdin/Grid'5000
- In both directions
  - Research → Production
  - Production → Research

First call to proposal in 2009

- Supported by C. Germain-Renaud and F. Desprez
- Total funding: 20,000 euros (10 from IdG, 10 from Aladdin)
- Lightweight procedure: scientific program on 3 pages
- Objective: establish consortiums and submit bigger proposals
- Selection ratio: 7/9

Selected Projects

- SimGlite, when SimGrid meets gLite
    - F. Suter – CC IN2P3
    - 5,000 euros
- Simulating Data-Intensive Applications
    - M. Quinson – LORIA/Nancy University
    - 5,000 euros
- Efficacité énergétique dans les grilles: de la recherche à la production
    - L. Lefevre – INRIA, LIP, ENS Lyon
    - 4,000 euros
- XWHEP : une grille de calcul globale securisee et interconnectee à EGEE
    - O. Lodygensky – LAL
    - 2,000 euros
- Criblage Virtuel de Semences
    - G. Da Costa – Uni. Toulouse
    - 2,000 euros
- Modélisations, Simulations et Calculs Hautes Performances pour l'énergie solaire
    - M. Daumas – ELIAUS
    - 1,000 euros
- Calcul à hautes performances sur processeurs GPU pour la biologie intégrative
    - D. Hill - Univ. Clermont
    - 1,000 euros

Next call in 2010

What is unchanged

- Lightweight procedure
- Calendar (at fall)
- Number of selected project (less than 10)

Some new propositions

- Fund a Master internship
- Setup a collaboration forum
  - As for European projects
  - Researchers can propose some ideas
  - Production can submit some problems

What Research Can Bring?

- A new middleware
- A better TCP protocol
- A High-Performance MPI
- An OS deployment solution
- A new programming paradigm
- A task scheduling algorithm
- Virtualization
- Energy savings
- A platform simulator
- A trustful emulator

What Production May Answer?

- A new middleware. Not if I have to rewrite all my codes
- A better TCP protocol. Get it integrated in Scientific Linux first
- A High-Performance MPI. Will it help my sequential jobs?
- An OS deployment solution. I may deploy twice a year at most
- A new programming paradigm. Fortran is just fine
- A task scheduling algorithm. Round robin works well for my workload
- Virtualization. only if it adds flexibility and reliability
- Energy savings. to compute more for less, but no resource shut down
- A platform simulator. I care only for my results, not how they were obtained
- A trustful emulator. Why would I slowdown my machines?

What Production Does Expect?

- Reliability, Reliability, Reliability
- Transparency, Transparency, Transparency
- Recovering gracefully of a failure
    - Better than trying to prevent it
- Improving an existing tool should be better perceived.
- Hot Topics
    - Large databases.
    - Virtualization (for administration comfort)
    - Interoperability
        - Production grids start to connect each other
    - Mastering the energy consumption
    - Preventing the anticipated crash into the memory wall
        - Memory does not grow linearly with the number of cores
        - This will become a problem for 1-proc jobs

What Production Can Bring?

- Realism!
- Real Applications (some with large societal impact)
- Real users, with concrete needs and expects
- Different use cases
  - Often harder than the comfortable ones we use in research
- A way to promote research results

Final Word

Research and production communities have to work hand by hand even though they look in opposite directions