



Enabling Grids for E-science

## **3<sup>rd</sup> EGEE User Forum Finance and Multimedia Session Summary**

*Sy Holsinger  
Trust-IT Services Ltd.*

*3<sup>rd</sup> EGEE User Forum (Clermont-Ferrand, France)  
11-14 February 2008*

[www.eu-egee.org](http://www.eu-egee.org)



- **Overview**

- Finance developing and advancing and now needs to be marketed, but from a user perspective
  - Companies know about Grid, but need to who is using it and how
  - Offer competitive market advantage
- Political or government streamlining
  - Learn from the past by understanding and analyzing life patterns
- Creative media moving towards:
  - Content based image retrieval
  - Storage, distributed tailoring and streaming of media files
  - Gaming

- **Presentations**

- Stock Analysis Application (Ezio Corso, ICTP/EU-IndiaGrid)
- Using Grid for mapping out taxation policies (Nikolaos Ploskas – Univ. Of Macedonia)
- “GridVideo”: a Grid-Based Multimedia Application (Giuseppe Iellamo – Univ. Of Messina )

- **Purpose**
  - Augment the Grid with Application-Level infrastructural services beyond core MW services
- **Why?**
  - To identify the building blocks in order to combine/provide solutions to other research problems going beyond monolithic applications
- **What?**
  - Latest trade volume, best-buy price, best-sell price, etc.
  - Whenever an event occurs such (i.e. new order arrival, etc.)
- **Analysis broken down into 2 Steps**
  - Characterize security's behaviour
    - *e.g. trade prices, returns, volatility, waiting time between trades, etc.*
    - *Resulting in 8TB of data*
  - Obtain output in form of tables and graphs
    - *MATLAB executable code in active development by researchers*

- **Why Grid?**

- Large Amounts of data

- overall 4 TB of input from 700 securities (stocks, bonds, options, etc.)
- Each security = several years data (1WN per security)

- Ideally suited to a Grid environment

- Large amount of data
- Embarrassingly parallel

- **Solution**

- gLite EGEE Grid (with WMS, WN, SE, etc.)

- 3 distinct parts:

- Business logic: code that encapsulates the finance/statistics knowledge
- Analysis interface: accepts analysis requests and manages GRID jobs
- Local environment manager: encapsulates GRID knowledge when interacting with Business Logic in WN.

- **Objective**
  - Achieve automatic processing of large numbers of files
  - Manage the business-logic engines in the WNs
  - Obtain full Grid Web Service
  - To be used by:
    - User-friendly web applications
    - Other web services for more elaborate operations
    - Directly by simple command-line clients
- **Potential exploitation of gLite in Finance**
  - Algorithm calibration
    - Something to be explored beyond this current application
- **Conclusions**
  - Attempt to put in practice the architectural view of achieving a solution through co-operating services
  - Will this serve as a true building block for researchers and eventually industry?
  - Aim to be up and running by June'08

- **Purpose**
  - Introduce an application which is a powerful tool for tracing taxation policy
  - Understand past patterns to then be used to change current policy
    - Allow a government to forecast its future surplus or deficit
- **Why Grid?**
  - Public Administration needs to have:
    - Reliable elements
    - Alternative models and scripts
  - Grid provides:
    - process of large volume of data
    - simultaneous control of different approaches
- **How?**
  - Creation of a model including multiple variables effecting tax policy or economy
    - the Net Interest Payments
    - the Government Purchases (G)

- **Example**
  - If a government's goal is to increase its surplus by 5%, it can create different approaches to achieve this goal, such as:
    - Increase only the taxes
    - Increase the taxes and decrease the Government Purchases
- **Issues**
  - Lack of real elements/data
  - Government slow response time
- **Conclusions**
  - A bigger sample of data leads to more accurate results
  - Forecasts can be more adequate
  - Could be effective tool for government
    - Gvt. could be able to see how it can change its taxation policy by understanding past patterns and comparing that to current policy
- **Future Actions**
  - Continue negotiations with government for real data
  - Once real data is obtained and pattern results achieved could then an analysis or reanalysis of current policy be done

- **What is Grid-Video?**
  - Grid computing paradigm for:
    - Storage
    - Distributed tailoring
    - Streaming of media files
- **How?**
  - 2 Steps
    - a media file is split, recorded on a SE
    - tailored by WN and the media file is streamed towards the user device
- **Uses**
  - gLite 3.0
    - Architecture sets up well for this kind of application
  - Mixture of Bash Shell and Java
    - JMS server: Apache ActiveMq.
- **Goal**
  - Simply to access multimedia contents on the Grid



- **Problems**

- If  $X$  is the number of chunks of the requested media
  - Submit  $X$  jobs each one recovers a media chunk, transcode it and stream it to the user device
- But that doesn't work:
  - If  $X$  jobs to the resource broker is requested, there is no guarantee on when these jobs will be scheduled
  - In order to enable the streaming the exact address of the WN used must be known
    - *this is not usually given by plain jobs*
  - Interactive jobs proved not to be a practical solution

- **Solutions**

- Java Message Service (JMS) because
  - Flexible communication service between various jobs
  - wide adoption and support
- Idle Jobs
  - is a sort of pre-submitted job that contains all the application logic but does nothing until it receives instruction via messages from the user interface

- **Advantages**
  - Latency
    - Lowered/Cut Down
      - *Idle Jobs method*
    - Improved latency
      - *Fibonacci Splitting*
        - Not same size, but progressive
  - Scalability
    - As long as there are Idle jobs, latency is not effected by an increase in number of users
- **Conclusions**
  - Porting application to Grid environments requires an important effort
  - Reliability is a great issue
    - A Grid is composed by heterogeneous system that changes over time
    - Non trivial multi job applications need some standardized way to communicate
  - Valuable example of how Grid could be used for the development of non-scientific applications

- **Overall session was interesting and lead to lively discussions**
- **Continue to market user case studies to attract industry**
  - Meet them
  - Inform them of Grid and EGEE
    - The benefits
  - Show them it exists
    - Grid monitor
  - Present them with User Case Studies
    - They will be more receptive if they see someone else is doing it
  - Currently gLite business case studies being collected
    - Use this in marketing
    - Presentations
    - One on one meetings, etc.