excelian

Opportunities for open source grid in the finance industry

Dr Adam Vile

Head of Grid, HPC and Technical Computing

.....approachable expertise

Grid in Finance – A Quick Summary

- Financial modelling continues to grow in complexity
  - More sophisticated models
    - some can take 30 mins to price on 100 node grid
  - More accuracy
    - More simulations and more points in the grid
  - More rapid pricing and risking
    - Intra-day and overnight
    - SLAs
    - Regulatory requirements
  - Higher volumes of instruments
    - example - 30,000 swaps in a single portfolio
- Continuing to drive the need for more and more compute

Extending the Definitions – Types of Grids In Finance

- Point solutions
  - Cluster type approach
  - Often directly connected with the business (of making money)
  - Usually addressing one specific problem
- Line of business grids
  - Owned and managed by a single LOB
  - Allows multiple job types
- Shared grids
  - May be shared between two or more business lines
  - Usually using lending and borrowing
- Enterprise grids
  - A single “compute backbone” that can service any request for compute according to agreed service levels
  - Often multi regional
  - Not multi entity

Growth of Enterprise Level Grids and the Introduction of Data Grid technology

- Enterprise grids require sharing of resources
- Current silos of 100s of nodes will be merged to form grids running multiple services with 1000s of nodes
- Sharing may be across regions and across and within departments
- **Prioritisation, Lending and Borrowing** and **Policy Based Allocation** are essential to facilitate sharing
- But sharing, and centralisation, is not popular with the business
  - SLAs have to be agreed
  - Loss of control
- It is, of course, very popular with infrastructure

- **The data problem** -
  - getting the right data to the right place at the right time!!!
  - Data sizes promise to be in the order of 4-8 GB per compute node
  - Market data changes intra second
  - Sharing of data, and calculation results is essential to improve efficiency
- **Caching or File system?**

How Do They Do It?

- With 3rd Party solutions from DataSynapse and Platform
  - More than 75% of grids in the city and on wall street are Datasynapse GridServer or Platform Symphony
- By writing their own solutions
  - Using DCOM, Sockets, SOAP,JINI
- By exploiting open source solutions
  - Condor
  - ICE
  - Grid Systems

Platform

Data Synapse

ZeroC

The Open Source Grid Initiative – Aims and Motivations

- Open source = free
  - The community supports itself
- Grid
  - coordinates resources that are not subject to centralized control
  - using standard, open, general-purpose protocols and interfaces
  - to deliver nontrivial qualities of service
  - "A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities." (Kesselman, Foster and Tuck 1998)
  - “coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations." (Kesselman and Foster 2000)
- Grid has been successfully used to enable the solution of computational problems that were unattainable within a single institution
  - i.e CERN (analysis of 10 Petabytes of data per year) uses the EGEE grid infrastructure **gLite**
    - 240 sites
    - 41,000 CPU’s
    - 45 countries
    - 30,000 jobs scheduled per day
- Why can’t we exploit this in the finance community?

Finance Industry Grid Requirements: Some Barriers to Adoption of Open Source?

- Lower latency, high throughput solutions
  - Stream as well as batch
  - Scheduling 10 jobs per second
- Job run time between 250ms to 3 hrs
  - Both Stateless and Stateful processing
  - Between KB and GB of data required
- Grids in excess of 10,000 compute nodes
  - with a utilisation of 70%
  - enterprise wide grids
  - charge-back capability
- Guaranteed completion
  - Failover and DR supported
  - “At most once” submission
  - Service levels agreed and achieved
- Supportable
  - Easy installation and maintenance
  - Available support and expertise
  - Grid is just middleware!!

Breaking into the Finance Industry Requires a Change in Frame

- Collaboration
  - *Open source grid has been designed to facilitate collaboration and share resources across boundaries*
  - **BUT** collaboration between companies is not a commercial reality
  - Open Source advantage - There may be a place for outsourced compute
- Productionisation
  - *Open source grid makes more resource available than could be owned by a single institution*
    - It is a clear business enabler
  - **BUT** deployment, resilience, DR, scalability and maintenance have large costs and must be addressed by any solution.
    - Evolutions of open source grid must be an "out of the box solution"
- Heterogeneity
  - *Open source grid installations support freely available operating systems and development tools*
  - **BUT** they must work across all environments – including windows clients
- Code base responsive to change
  - *Open source grid has often been developed on a best efforts basis, around medium term time-scales*
  - **BUT** companies need a codebase that can can respond to new requirements quickly, and a support network to facilitate these changes

Rather than trying to change the fundamental vision of open source grid, perhaps we can focus on new requirements, new visions, and new potential customers that may open the door

- Utility compute
- Gridonomics *
- Interoperability

*Term coined by the 451 group

Utility compute

- Pay as you go
  - Not everyone can afford to maintain their own grid resources
  - Capitalising on the current cpu/hr offerings
- Potential for ASP solutions offered to smaller institutions
  - Not suitable for intra-day calculations
  - Could offer DR solution
- Not only compute but also Data

Gridonomics

- How do we charge for compute in a shared resource pool?
  - fully internal
  - fully external
- Cost models
  - Pay per cpu/hr
  - Commodity model (could lead to derivatives)
  - Do we all pay the same or is the model based upon priority, resource type etc...

Superscheduling and Grid Interoperability

- Interoperating
  - between various grid silos
  - between vendors
  - between business areas
  - between “enterprise” grids
  - internally and externally
- Super scheduling
  - Needs to ensure the same guarantees as the standard scheduler
  - How to manage deployment and resources
- A common grid API?
  - what is wrong with existing attempts?

Opportunities for gLite in Finance

- Batch not stream
  - A focus on batch processing for smaller companies
- Pay as you go
  - Infrastructure to support an ASP solution?
- Interoperability
  - Bringing together disparate grid Islands into an atoll* through super scheduling
- Data grid/data cache
  - integration with the major data grid solutions
- A common grid API
  - Supports interoperability, data movement, and super scheduling

*This metaphor was suggested to me by Ian Osborn from “Grid computing now”