

# Grid in Financial Services and Science: a comparison

Peter Jenkins

CSC Finland

[peter.jenkins@csc.fi](mailto:peter.jenkins@csc.fi)

CSC – Tieteen tietotekniikan keskus Oy  
CSC – IT Center for Science Ltd.



I'm from there

CSC

You are here!

# Grid with a chance of clouds



- High level comparison Science and Finance
- Finance is big too
  - Several tier 1 banks have  $\geq 30k$  cores
  - Not on top 500
- Cloud is hard for both science and finance
  - Different reasons
- Immature technology and lack of skills
- My background

# Outline

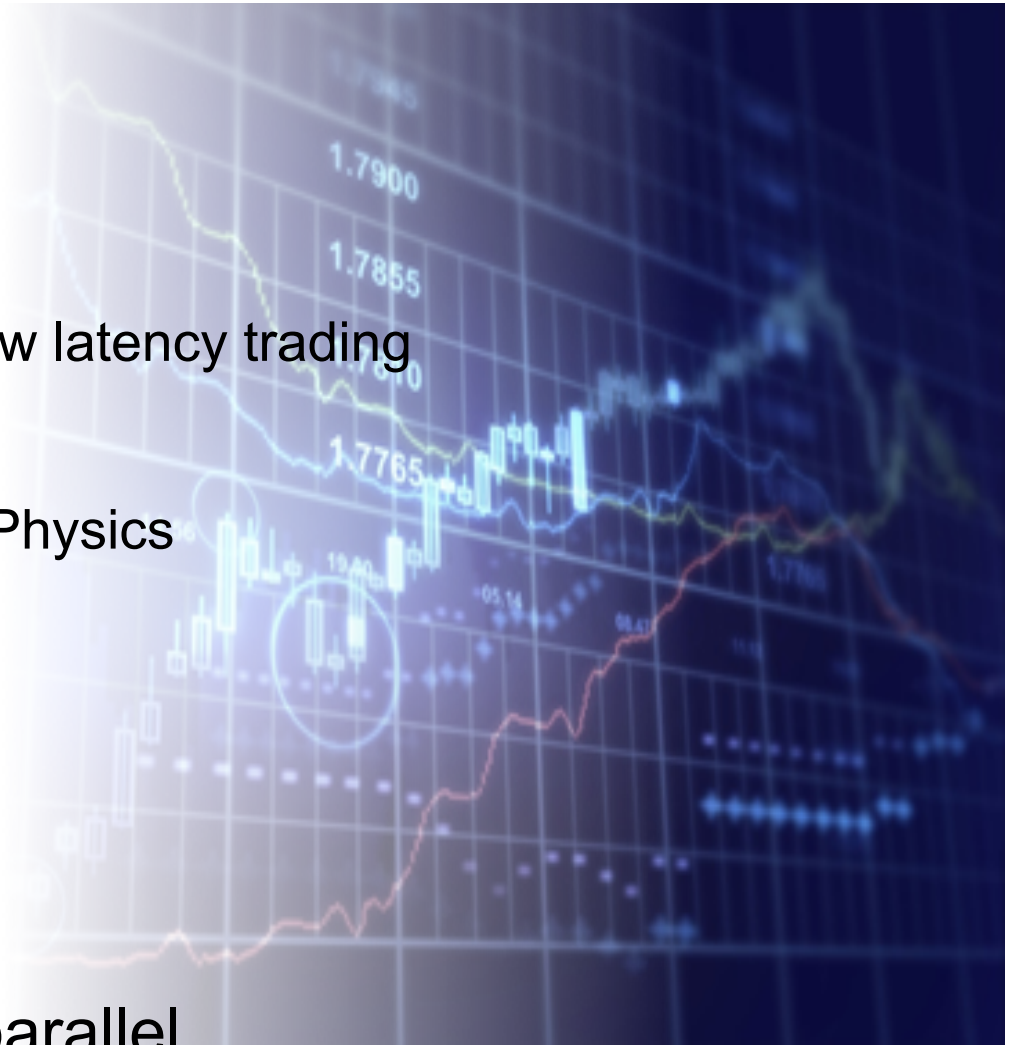


- Applications
- Middleware
- Resource utilisation
- Virtualisation
- Cloud

# Applications



- Typical applications
  - Finance
    - Risk analysis, Pricing, Low latency trading
  - Sciences
    - Biosciences, Chemistry, Physics
- Requirements
  - Time scales
  - Data
  - Memory
  - Finance: Embarisingly parallel



# Middleware: Finance



- Costs
- Notable features
  - Web based interfaces
  - Redundancy
  - Programming model
  - Middleware integration – Data Caching
- Requirements
  - Job submission speed/throughput
  - Reliability
  - APIs (Client and Worker)
- Products



# Middleware: Science

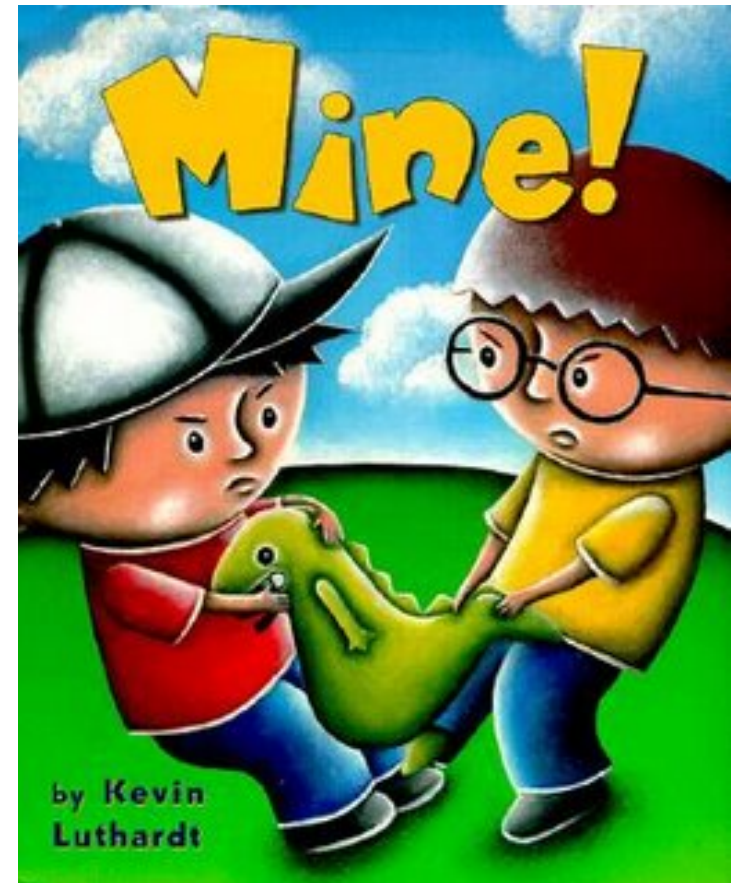


- Costs
  - vary by industry
  - free software can be ok
- Features
  - Mature (old) products
  - Batch job systems - command line interfaces
  - MPI integration
  - Integration with external grids/systems
- Requirements
  - Fair sharing/utilisation and accounting, Scale
- Products
  - Condor, Platform LSF, Oracle Grid Engine, Slurm

# Resource sharing



- Similar state of maturity in both
- Desire for enterprise grid
  - Now called internal cloud
- Finance
  - Flexibility of infrastructure a key asset: hard to achieve
- Why?
  - People and organisations
  - Legacy applications
  - Wide mix of environments



# Virtualisation



- Limited usage today in Grid HPC. Why?
- Perception problem?
- Technology limitations?
  - Additional layer of complexity
  - Counter intuitive to horizontal scale?
  - Hard to manage 1000's physical hosts?
  - Science only
    - IO performance
    - Infiniband



# Cloud: Definition for here



- What I've actually seen
- External cloud
  - Limited adoption but for very different reasons
- Here "cloud" means IaaS - Amazon EC2 etc
  - Infrastructure as a service
- IaaS = managed hosting of VMs with an API
- PaaS out of scope but interesting
  - Google App engine
  - Higher level domain specific services



# Cloud: Finance



- Reliability
  - expensive and complex multi vendor/multi site solutions
- Costs
  - Internal vs. External, Measurements hard, conducted internally
- Data protection and compliance
  - physical location of data
- Application latency concerns
  - mainly outmoded thinking
- Perceived security risk (outmoded thinking)
- Security
  - Regulatory compliance and in theory, strict governance
  - shortage and technology maturity limits flexibility

# Cloud: Science



- Virtualisation issues
- Limitations of service providers
  - MPI
  - Parallel filesystems
  - better fit for internet/web 2.0 (perception?)
- Large data sets
- Central/Project funding bad fit for pay as you go?
  - Virtual infrastructure gateway setup required
  - Access control challenge
- Skills shortage
- Collaborative yet secretive ambiguity nature of research
  - Access control challenge

# Conclusions



- Both Finance and Science Grid/HPC is complex
- Clouds don't make this easier
- Greenfield application specific PaaS type cloud?
  - Will Google Spreadsheets auto parallelised computation?
    - Is Excel on the grid really bad?
  - Will Matlab one day have cloud compute offload?
- Finance can't do this - will commercial IaaS improve enough?
  - Grid/HPC tailored virtual appliances
    - GigaSpaces XAP
  - Finance will continue to rewrite/rearchitect applications to remain competitive

# Summary



- Applications
- Middleware
- Resource utilisation
- Virtualisation
- Cloud

# Extra slides



# Hardware



- Broadly similar servers
  - Blades
- Networks
  - science applications sometimes requires "capability" machines
  - Fatter interconnects than standard Ethernet or InfiniBand can provide
  - Rarely genuine SMP
  - Finance is mainly Ethernet
    - High frequency/low latency trading use InfiniBand

# Storage: Finance



- Include data with job
- NFS or CIFS if at all
  - No HPC specific file systems
- databases often involved
  - scalability problem
  - Biosciences already have this problem
- In memory caches
  - Programming model integration
  - Gigaspaces
  - Oracle Coherence
  - Gemfire Gemstone

# Storage: Science



- NAS typically assumed
  - application requirement
  - interprocess communication
  - program distribution
  - configuration management
  - user program logging
- Large data sets often mandate parallel filesystems

# Reliability



- multiple sites
  - multiple geographies
  - DR site active-active
- Multiple environments
  - Strict change management
- grid scheduler is critical infrastructure
  - "master" always in some way HA
  - Shared filesystems not used for DataSynapse
- middleware handles failures/retrys