Heterogeneous high-performance requirements for parallel applications
The hardware and software requirements for parallel applications depend on the problem size, type and the number particles / parameters, the degree of parallelization possible, the load balancing over different processors / memory, the calculation type and the input / output and storage amount. These determine the suitability of a hardware and software solution optimizing the hardware use. Computational time is the meta-parameter for optimization. The interaction with users, the management of jobs, accounting, and billing, are getting increasingly important and make further optimizations necessary. Thus, parallel high-performance grid computing has to combine the challenges of optimizing a dedicated problem within a heterogeneous hardware setting with the right management. By combination of small grid jobs with huge parallel jobs, CPU/core usage efficiencies > 99.9% can be reached.

DNA/nucleosome MD simulations on a parallel high-performance grid

SMD simulation of a 44 base pair long DNA segment over 50 ps. The complete simulation was performed in a water box with ions, to neutralize the charges of the DNA (not visualized). In the middle of the DNA, an atom was selected and a pulling force (to the right) was applied resulting in bending. A) Nucleosome in a box with water molecules and ions in a VDW visualization. B) Start structure of a nucleosome for an MD simulation (water/ions not shown). The DNA is wrapped around the histone core. C) Same nucleosome as in B) but after 20 ns of simulation. Changes in the DNA and histone tails are clearly visible.

Management of users and jobs as well as accounting and billing

Speedup

The maximum resource efficiency in parallel high-performance infrastructures can be reached by advanced user management. This involves information about the grid resources and their technical capabilities for parallel jobs, education on parallelization in general, the simultaneous optimization of applications towards several dedicated hardware, and technical and psychological training how trivial and nontrivial parallel jobs run in the grid and how this influences the overall meta-performance of the system. The accounting and billing in parallel high-performance grids also needs new strategies, due to the information aggregation over different cores, and the more complex accounting structure. During the waiting time for the necessary amount of cores, small parallel or trivial filler jobs can be executed. This is very similar but more complex than for grid visualization approaches. For billing, this suggests a variety of opportunities, where the price can be coupled to the accounting via an integrative model with weighted scenarios. Vice versa also the pricing can be used to put pressure on the user to optimize his application concerning the provider goal to run the resource at the optimum.

Conclusion

In the life-science and health-care sectors demanding IT and especially large scale grid resources are needed to analyze huge and complex systems. We showed here with MD simulations of different sizes, how the presence of large grid clusters within grid infrastructures allows this parallel high-performance grid computing efficiently and thus combines the benefits of dedicated super-computing centres and grid infrastructures. Therefore, accounting and billing can be used to motivate the users to use the hardware in a most efficient way, which not always results in shortest computation times. Thus, we combined classic with parallel high-performance grid usage, while increasing the efficiency of IT resource providers. This is beyond the mere "yes-we-can" a great opportunity for the life-science and health-care sectors and allows reaching the maximum resource efficiency.
Parallel High-Performance Grid Computing: Capabilities and Opportunities of a Novel Demanding Service and Business Class allowing Highest Resource Efficiency

Kepper, N., Ettig, R., Dickmann, F., Stehr, R., Grosveld, F. G., Wedemann, G. & Knoch, T. A.


Abstract

The hardware and software requirements for parallel applications depend on the problem size, type and the number particles / parameters, the degree of parallelization possible, the load balancing over different processors / memory, the calculation type and the input / output and storage amount. These determine the suitability of a hardware and software solution optimizing the hardware use. Computational time is the meta-parameter for optimization. The interaction with users, the management of jobs, accounting, and billing, are getting increasingly important and make further optimizations necessary. Thus, parallel high-performance grid computing has to combine the challenges of optimizing a dedicated problem within a heterogeneous hardware setting with the right management. By combination of small grid jobs with huge parallel jobs, CPU/core usage efficiencies > 99.9% can be reached. In the life-science and health-care sectors demanding IT and especially large scale grid resources are needed to analyse huge and complex systems. We showed here with MD simulations of different sizes, how the presence of large grid clusters within grid infrastructures allows now parallel high-performance grid computing efficiently and thus combines the benefits of dedicated supercomputing centres and grid infrastructures. Therefore, accounting and billing can be used to motivate the users to use the hardware in a most efficient way, which not always results in shortest computation times. Thus, we combined classic with parallel high-performance grid usage, while increasing the efficiency of IT resource providers. This is beyond the mere “yes-we-can” a great opportunity for the life-science and health-care sectors and allows reaching the maximum resource efficiency.

Corresponding author email contact: TA.Knoch@taknoch.org

Keywords:

Genome, genomics, genome organization, genome architecture, structural sequencing, architectural sequencing, systems genomics, coevolution, holistic genetics, genome mechanics, genome function, genetics, gene regulation, replication, transcription, repair, homologous recombination, simultaneous co-transfection, cell division, mitosis, metaphase, interphase, cell nucleus, nuclear structure, nuclear organization, chromatin density distribution, nuclear morphology, chromosome territories, subchromosomal domains, chromatin loop aggregates, chromatin rosettes, chromatin loops, chromatin fibre, chromatin density, persistence length, spatial distance measurement, histones, H1.0, H2A, H2B, H3, H4, mH2A1.2, DNA sequence, complete sequenced genomes, molecular transport, obstructed diffusion, anomalous diffusion, percolation, long-range correlations, fractal analysis, scaling analysis, exact yard-stick dimension, box-counting dimension, lacunarity dimension, local nuclear dimension, nuclear diffuseness, parallel super computing, grid computing, volunteer computing, Brownian Dynamics, Monte Carlo, fluorescence in situ hybridization, confocal laser scanning microscopy, fluorescence correlation spectroscopy, super resolution microscopy, spatial precision distance microscopy, auto-
fluorescent proteins, CFP, GFP, YFP, DsRed, fusion protein, in vivo labelling, information browser, visual data base access, holistic viewing system, integrative data management, extreme visualization, three-dimensional virtual environment, virtual paper tool.

**Literature References**


