

Стіренко С.Г.¹, Кушніренко Н.В.¹, Хондар В.С.²

¹НТУУ «КПІ» ФІОТ; ²ННК «ПСА» НТУУ «КПІ»

Засоби підвищення ефективності проходження задач у GRID середовищі

Сучасні тенденції в області високопродуктивних обчислень розвиваються у сторону підвищення вимог додатків до обчислювальних ресурсів. Нарощування потужностей, особливо у несприятливі періоди розвитку економіки, не дають можливості оптимально розв'язувати сучасні задачі. Як наслідок, постійно зростають витрати на апаратне та програмне забезпечення, також збільшуються і площі обчислювальних центрів.

З іншого боку, моніторинг ресурсів учбових комп'ютерних класів показує досить низьке завантаження окремих комп'ютерів, які мають досить високу локальну продуктивність. Таким чином, використовуючи технологію Grid Computing [1] з'являється можливість об'єднати та керувати розподіленими обчислювальними ресурсами, системами зберігання даних та каналами передачі даних.

Але, окрім вже існуючих засобів створення Grid-інфраструктур [2], лишається не розв'язаним набір важливих наукових задач, які втримують широке використання цієї технології. Ефективне керування ресурсами в розподіленому середовищі – одна з таких задач.

Розробка спеціальних засобів планування, які забезпечать керування потоком задач в гетерогенному середовищі, значно підвищить ефективність використання усієї Grid-інфраструктури. Дослідження в цій області активно проводяться (наприклад [3,4]), але застосування подібних розв'язків значно звужує область їх використання, а також накладає істотні обмеження на кількість одночасно існуючих елементів у Grid-системі та потребують від користувача знання спеціальних мов програмування.

Потік задач ("workload") може бути описаним на основі статистичних даних використання вже існуючої Grid-системи. Для опису таких задач використовується текстовий файл у спеціальному форматі [5]. Окремі рядки цього файлу містять інформацію про властивості задачі, час її породження та роботи. Синтетичний потік задач дозволяє визначити в моделі потік задач по наступним початковим параметрам: властивостям задачі; періоду – часу до породження чергової задачі; кількості породжуваних задач; затримки перед початком породження першої задачі. При цьому користувач повинен мати можливість задати визначення свого власного потоку задач.

Система планування повинна містити в собі готові шаблони для відображення завантаженості системи – загальної та з розбиттям по окремим кластерам, часу очікування задачі в черзі – середнього та пікового з розбиттям по класам задач, властивості пропускання, як по кількості задач, так і з використанням інтегральної оцінки.

Література

1. Foster I., Kesselman C. The GRID: Blueprint for a New Computing Infrastructure. Second edition, Morgan Kaufmann Publishers, 2004.
2. The microgrid: Using emulation to predict application performance in diverse grid network environments/ H.Xia, H.Dail, H.Casanova, A.Chien// In Proceedings of the Workshop on Challenges of Large Applications in Distributed Environments (CLADE'04), IEEE Press, 2004.
3. Legrand A., Marchal L., Casanova H. Scheduling distributed applications: The simgrid simulation framework // Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid 2003 (CCGrid2003), 2003, Pp. 138–145.
4. Parallel workloads archive <http://www.cs.huji.ac.il/labs/parallel/workload/>