

УДК 004.27

Тиртишніков. О.І., к.т.н, доцент

Мільченко Т.О., студент

*Полтавський національний технічний
університет імені Юрія Кондратюка*

НАВЧАЛЬНА ПАРАЛЕЛЬНА ОБЧИСЛЮВАЛЬНА СИСТЕМА: МОЖЛИВОСТІ ПОБУДОВИ ТА ВІРІАНТИ КОНФІГУРАЦІЇ

Для вивчення архітектури та методів програмування комп'ютерних систем вкрай бажаною є наявність реальних навчальних паралельних обчислювальних систем, що мають декілька процесорів та розподілену пам'ять. При цьому, наприклад, продуктивність таких систем може бути невисокою. В роботі досліджуються можливості побудови навчального кластера на основі доступного апаратного та програмного забезпечення з метою обґрунтування вибору оптимального за критерієм ефективність / вартість варіанта навчальної паралельної обчислювальної системи.

***Ключові слова:** кластерна система, суперкомп'ютер, обчислювальний вузол, паралельне програмування.*

Вступ

В останні роки відбувся суттєвий прогрес у суперкомп'ютерних технологіях та проникнення цих технологій у все нові сфери людської діяльності. Суперкомп'ютер став потужним інструментом, який дозволяє форсувати просування науково-технічної думки в багатьох галузях. Провідні країни світу створили та використовують цей інструмент для вирішення особливо складних задач науки, освіти, економіки, для створення довгострокових прогнозів та забезпечення національної безпеки.

Кластером називається паралельна обчислювальна система з розподіленою пам'яттю, що побудована з компонентів загального призначення, з єдиною точкою доступу, однорідними обчислювальними вузлами і спеціалізованою мережею, що забезпечує ефективний обмін даними [1, 2, 5].

Стандартними методами досягнення високої продуктивності є застосування вузькоспеціалізованих процесорів цифрової обробки, підвищення продуктивності серійних процесорів, розпаралелювання задачі на множині процесорів та об'єднаних у мережу систем. Вибір конкретної реалізації кластерної системи залежить від особливостей вирішуваної задачі і наявних ресурсів [1, 2, 5].

Кластерна система може проектуватися під яку-небудь спеціальну задачу, наприклад, для розрахунків на міцність (кластер спеціального призначення) або для виконання широкого класу паралельних програм (кластер загального призначення). Вона застосовується там, де необхідно істотно підвищити надійність (коефіцієнт готовності) і отримати значну обчислювальну потужність шляхом розпаралелювання задач.

Для вивчення архітектури та методів програмування комп'ютерних систем вкрай бажаною є наявність реальних навчальних паралельних обчислювальних систем, що мають декілька процесорів та розподілену пам'ять. При цьому, наприклад, продуктивність таких систем може бути невисокою.

Апаратний спосіб реалізації

У секторі "академічних" суперкомп'ютерів, що належать університетам, зазвичай використовуються найдешевші рішення, що складаються з набору ПК, з'єднаних Ethernet-мережею і працюють під управлінням безкоштовних операційних систем на зразок Linux.

Основні проектувальні параметри кластера:

– параметри обчислювального вузла і керуючого комп'ютера: тип і розрядність процесорів, кількість процесорів, тип та об'єм зовнішніх пристроїв

зберігання даних, об'єм кеш-пам'яті, тип та об'єм внутрішньої пам'яті, продуктивність внутрішньої шини даних, тип зовнішніх інтерфейсів;

– кількість обчислювальних вузлів;

– параметри обчислювальної мережі: латентність, пропускна спроможність, протокол обміну, топологія;

– пропускна спроможність керуючої мережі [1, 2, 5].

Обчислювальні вузли – це комп'ютери кластерної мережі, які складають основу її обчислювальної потужності. На них встановлюється спеціальне програмне забезпечення, яке дозволяє користувачам запускати свої завдання і контролювати їх стан. Кожний вузол слугує для того, щоб одна або більше задач виконувались як частина загального рішення. Обчислювальні вузли спільно використовують розподілені ресурси і серверне програмне забезпечення. Клієнти підключаються до кластера, не знаючи про те, який саме комп'ютер буде насправді займатися їх обслуговуванням. При штатній роботі кластера всі ресурси розподілені по вузлах і активовані. У кожний момент конкретний ресурс керується одним із вузлів кластера, – яким саме, визначено у конфігурації ресурсу або задано у процесі роботи адміністратором.

Безперебійність доступу, що забезпечується кластером, досягається за рахунок своєчасного виявлення порушень у роботі апаратних і програмних засобів і автоматичного переносу процесів обробки даних на справний вузол. У випадку відмови вузла або ресурсів система передає частину ресурсів на інший вузол і забезпечує їх доступність клієнтам.

Бажано, щоб обчислювальні вузли були однаковими комп'ютерами від одного виробника. Кластер може використовувати переваги сучасних, широко розповсюджених платформ персонального комп'ютера. Обчислювальними вузлами можуть бути як однопроцесорні, так і багатопроцесорні (SMP) комп'ютери, побудовані на базі обчислювальних компонентів загального призначення.

При комплектації обчислювальних вузлів кластера цілком можливо відмовитися від жорстких дисків. У бездисккових вузлів є низка переваг – менше

енергоспоживання і тепловиділення, нижче «гучність», вище відмовостійкість. Однак, висока продуктивність кластера природнім чином призводить до того, що на ньому буде оброблюватись більша кількість даних – отже, їх прийдеться десь зберігати. Для цих цілей можна використати сховище даних необхідного об'єму, підключене до керуючого комп'ютера. Таке рішення виявиться більш коштовним, а дані, які постійно будуть передаватися мережею, зменшать її пропускну спроможність. У рамках бюджетного кластера, коли мережа використовується і для обміну даними між паралельними процесами, і для передачі даних, використання центрального сховища може негативно відбитися на обчислювальній продуктивності кластера у цілому. Тому слід укомплектувати кожний вузол невеликим жорстким диском. За необхідності можна об'єднати всі диски в одне розподілене сховище, встановивши кластерну файлову систему.

Кожний вузол кластера повинен мати два мережевих адаптери. Один – для підключення до загальної мережі, другий – для обміну даними між вузлами кластера. Якщо обчислювальні вузли не будуть використовуватись як робочі місця користувачів, немає необхідності встановлювати на них відеокарти і монітори. Клавіатури і миші на обчислювальних вузлах також не потрібні.

Серед вузлів кластера окремо виділяють керуючий комп'ютер, на який встановлюється керуюче програмне забезпечення. Дана програмна складова надає користувачам інтерфейс доступу до кластера і здійснює керування обчислювальними вузлами, посылаючи на них команди і запити. Таким чином, керування кластером виконується як єдиною обчислювальною системою. Керуючий комп'ютер забезпечує керування всіма вузлами кластера за допомогою керуючої мережі. Він виконує додаткові функції файлового сервера, є станцією резервного зберігання даних. Через його консоль, або віддалено через засоби безпечного доступу до нього, здійснюється запуск задач на кластері. Керуюча консоль дозволяє адміністратору контролювати стан усіх ресурсів кластера. Можна переносити робоче навантаження з одного комп'ютера на інший у межах кластера. Така можливість відкриває шлях до

рівномірного розподілення обчислювального навантаження між вузлами кластера, а також вивільняє один із комп'ютерів кластера, наприклад, для проведення планових профілактичних робіт. Використовувати керуючий комп'ютер як обчислювальний вузол не можна, адже робота користувачів буде ускладнена.

Функції керуючого комп'ютера можуть бути реалізовані або через одну, або через групу робочих станцій. Для прискорення роботи локальних додатків слід установити на даному комп'ютері швидкий жорсткий диск або RAID-контролер. До процесора керуючого комп'ютера великих вимог не пред'являється тому що обчислювальні задачі на ньому не виконуються, а для компіляції великої потужності не потрібно. Якщо заздалегідь відомо, що користувачів буде багато, то має сенс використовувати комп'ютер із декількома процесорами.

За інших рівних умов для кластера слід вибирати комп'ютери з кеш- та основною пам'яттю, що мають найбільші обсяг та швидкодію, оскільки ці характеристики істотно впливають на час виконання паралельної програми.

Обчислювальна мережа кластера може бути реалізована за допомогою наступних мережевих технологій: Quadrics, Myrinet, SCI і інших [1, 2].

Керуюча мережа кластера з протоколом TCP/IP об'єднує вузли кластерного рівня у звичайну локальну мережу (LAN) [1, 2]. Вона реалізується з використанням широко розповсюдженої технології Ethernet як 100 Мбіт/с (100 Base-T), так і 1 Гбіт/с (1000 Base-T), і призначена для керування кластерною системою, підключення робочих місць користувачів, інтеграції суперкомп'ютера у локальну або глобальну мережу.

При побудові кластерів особливу увагу приділяють системам їх розміщення, охолодження і електроживлення. При 100% навантаженні активна потужність, що споживається одним вузлом кластера, складає 90-120 Вт і розсіюється у вигляді тепла. Арифметичні підрахунки показують, що кластер із 16 вузлів зрівняється за потужністю з побутовим нагрівачем. Наприклад, кластер "СКІФ К-1000" споживає більше 89 кВт енергії, з якої практично вся

йде у тепло. Такої потужності було б достатньо для обігріву невеличкого дому, тому потрібен продуманий тепловий дизайн. У сучасних кластерних суперкомп'ютерах застосовують повітряне охолодження, і необхідний температурний режим забезпечується двома варіантами реалізації. Перший – це продуманий тепловий дизайн обчислювального вузла: стандартні шасі модернізуються таким чином, щоб повітряний потік, створюваний внутрішніми вентиляторами, максимально ефективно охолоджував процесори. Другий – підтримка робочої температури у приміщенні: гаряче повітря повинно бути або відведено від вузлів і кондиціоновано, або спрямовано за межі приміщення.

Може знадобитися встановити у приміщенні, де розміщено кластер, дезволожувач або фільтр повітря. Висока вологість або пил, які у звичайній ситуації виведуть з ладу один комп'ютер, у приміщенні, де знаходиться кластер, можуть спричинити значний збиток.

Електроживлення кластера повинно бути безперебійним. Його забезпечення можна організувати за допомогою матричного джерела безперебійного живлення (ДБЖ), що обслуговує весь кластер, або за допомогою окремих ДБЖ, до яких підключено по декілька вузлів. Установлювати матричне ДБЖ економічно вигідно тільки при достатньо великій кількості вузлів. Розраховуючи потужність ДБЖ, слід визначити, протягом якого часу кластер повинен працювати у випадку зникнення напруги. Для підтримки роботи чотирьох вузлів протягом 15-20 хвилин повинно вистачити ДБЖ потужністю 1000 ВА. Вибираючи ДБЖ, слід подбати і про те, щоб він вчасно міг попереджувати операційну систему про відсутність напруги живлення у мережі, щоб система змогла коректно завершити всі задачі, перш ніж вичерпається енергія акумуляторів ДБЖ, якщо напруга так і не з'явиться.

Програмний спосіб реалізації

Для кластера можна вибрати практично будь-яку операційну систему. Традиційно для створення кластерів використовується Linux (більш 70% систем

Тор 500) або інші різновиди Unix (решта 30%).

Після етапу проектування здійснюється інтеграція обраного устаткування відповідно до компоновочного рішення. Після цього встановлюється системне і прикладне програмне забезпечення, яке дозволяє використовувати дане устаткування як єдину обчислювальну кластерну систему.

Системне програмне забезпечення кластера включає у себе: операційну систему вузлів, систему керування кластером – засоби, що забезпечують роботу з кластером як із єдиною обчислювальною системою, паралельне оточення кластера (MPI, PVM). Операційна система кластера слугує для керування всіма функціями кластера. Всі вузли кластера працюють під керуванням власної копії операційної системи. Її вибір ґрунтується на рекомендаціях розробників програмного забезпечення. Кластер, як правило, працює під керуванням одного з різновидів операційної системи – Unix – багатокористувацької, багатозадачної мережевої операційної системи, зокрема є кращим використання операційної системи Linux (Red Hat Enterprise Linux, SuSe Linux Enterprise Server, Scientific Linux) для кластера. Linux є однією із самих надійних, ефективних і перспективних операційних систем, яку сьогодні більшість комерційних і державних організацій вибирають базовою для додатків і перспективних розробок у галузі паралельних обчислень. Під керуванням Linux доступна більша частина серверного програмного забезпечення, компіляторів, бібліотек, засобів налагодження, для багатьох програм є вихідні коди і численна документація. Linux розповсюджується вільно (безкоштовно) з вихідними текстами.

Для забезпечення можливості ефективної експлуатації кластера як єдиного обчислювального комплексу використовується система керування кластером (Cluster Management System, CMS). На керуючому комп'ютері кластера встановлюється програмне забезпечення CMS, яке керує запуском програм на кластері, планує виконання пакетних завдань, а також їх розподілення між обчислювальними вузлами кластера. Адміністратор обчислювального кластера і диспетчер завдань – це додатки, які призначені для адміністрування кластера і

керування завданнями користувача. Ці програмні засоби керують розподіленням навантаження і збільшують ефективність роботи обчислювальної системи. Адміністратор обчислювального кластера слугує для конфігурації кластера, а також моніторингу стану обчислювальних вузлів і кластера у цілому. За допомогою диспетчера завдань користувачі кластера здійснюють створення, відправлення і відслідковування стану своїх завдань. Найбільш розповсюджені пакети адміністрування компаній Altair, Platform, Sun і Scali. Однією з найбільш популярних систем керування кластером є вільно розповсюджувана система Oscar, розробка якої підтримується Gelato Community.

Паралельним оточенням обчислювального кластера є система організації паралельних кластерних обчислень MPI (Message Passing Interface) [3, 4]. Серед вільно розповсюджуваних реалізацій MPI широко використовуються MPICH, LAM.

Паралельні програми для кластерних систем розробляються з використанням послідовних мов програмування (C/C++, Fortran) і комунікаційних бібліотек MPI, PVM. Можлива інсталяція компіляторів як вільно розповсюджуваних (GCC), так і комерційних (Intel, Portland Group).

Прикладне програмне забезпечення кластера звичайно включає наступні основні додатки:

- компілятори (Fortran, C/C++, JDK, ADA, Modula-3);
- паралельні бібліотеки високого рівня (ATLAS, SCALAPACK);
- засоби автоматичного і полуавтоматичного розпаралелювання послідовних програм (BERT 77, FORGE, KAP, PIPS, VAST);
- прикладні паралельні програми;
- бази даних (Oracle, SQL Server);
- системи керування ресурсами підприємства (ERP);
- засоби обробки повідомлень і поштові системи;
- системи взаємодії з клієнтами (CRM).

Висока продуктивність кластера і зведення до мінімуму часу простоїв додатків досягається завдяки тому, що:

- у випадку збою програмного забезпечення на одному з вузлів додаток продовжує функціонувати або автоматично перезапускається на інших вузлах;
- вихід із ладу одного або декількох вузлів не призведе до краху всієї кластерної системи;
- профілактичні і ремонтні роботи, реконфігурацію або зміну версій програмного забезпечення, як правило, можна здійснювати у вузлах кластера по черзі, не перериваючи роботи інших вузлів.

Серед способів задіяти обчислювальні потужності кластера можна виділити наступні:

- запускати багато однопроцесорних задач, якщо необхідно провести багато незалежних обчислювальних експериментів із різними вихідними даними;
- запускати готові паралельні програми (для деяких задач доступні безкоштовні або комерційні паралельні програми, які за необхідності можна використати на кластері);
- викликати у програмах паралельні бібліотеки (для деяких областей, таких як лінійна алгебра, доступні бібліотеки, що дозволяють вирішити широке коло стандартних підзадач із використанням можливостей паралельної обробки);
- створювати власні паралельні програми шляхом вставки паралельних конструкцій у наявні паралельні програми.

Для прискорення обробки даних на кластері необхідно максимально прискорити обчислення на одному процесорі, для чого можна вжити наступних заходів:

- підібрати необхідні опції оптимізації компілятора;
- використати оптимізовані бібліотеки (якщо деякі стандартні дії, такі як множення матриць, займають значну долю часу роботи програми, то має сенс використати готові оптимізовані процедури, що виконують ці дії, а не програмувати їх самостійно);

– виключити автоматичне скидання даних із пам'яті на диск (кожний процес повинен зберігати не більше даних, ніж для нього доступно оперативної пам'яті);

– більш ефективно використовувати кеш-пам'ять (у випадку можливості змінювати послідовність дій програми, потрібно модифікувати програму так, щоб дії над тими самими або підряд розміщеними даними виконувались також підряд);

– більш раціонально працювати з тимчасовими файлами (якщо програма створює тимчасові файли у поточному каталозі, то необхідно перейти на використання локальних дисків на вузлах);

– використовувати найбільш доцільні типи даних (у деяких випадках замість 64-розрядних чисел із плаваючою крапкою подвійної точності може бути доцільним використання 32-розрядних чисел одинарної точності або, навіть, цілих чисел).

Висновок

Аналіз продуктивності, вартості і співвідношення “ціна–продуктивність” для обчислювальних систем різної структури показав, що кластери мають суттєві переваги перед мейнфреймами і серверами на базі RISC-архітектури.

Запропоновані підходи і принципи дозволяють створювати кластерні системи, що відповідають наступним вимогам:

– зменшують витрати на адміністрування локальної мережі, тобто мають гарну керованість, дозволяють гнучко змінювати кількість і склад апаратних засобів і програмного забезпечення відповідно до мінливих вимог задач, що вирішуються; – забезпечують високу готовність – можливість дуже швидко перерозподілити роботу на інші процесори всередині кластера при відмові одного процесора;

– забезпечують мобільність програмного забезпечення – можливість запуску програмних систем на різних апаратних платформах, а також

гарантують можливість застосування однакових людиномашинних інтерфейсів на всіх комп'ютерах, що входять у кластерну мережу;

– автоматично розподіляють навантаження між вузлами кластера, у результаті чого один додаток може працювати на декількох вузлах і використовувати їх обчислювальні ресурси, що призводить до ефективного використання ресурсів кластера і підвищення його швидкодії.

Посилання

1. *Воеводин Вл. В., Жуматий С. А. Вычислительное дело и кластерные системы. – М.: Изд-во МГУ, 2007. – 150 с.*
2. *Шпаковский Г. И., Верхотуров А.Е., Серикова Н. В. Руководство по работе на вычислительном кластере: Пособие для студентов естественно-научных специальностей. – Минск: БГУ, 2004. – 171 с.*
3. *Шпаковский Г. И., Серикова Н. В. Программирование для многопроцессорных систем в стандарте MPI. – Мн.: БГУ, 2002. – 323 с.*
4. *Воеводин В. В., Воеводин Вл. В. Параллельные вычисления. – СПб.: БХВ-Петербург, 2002. – 608 с.*
5. *Орлов С.А., Цилькер Б.Я. Организация ЭВМ и систем: Учебник для вузов. 2-е изд. – СПб.: Питер, 2011. – 688 с.*

Рецензент: Мавріна М.О., к.т.н., доцент.

Authors: Tyrtysnikov O.I., Milchenko T.O.

EDUCATIONAL PARALLEL COMPUTING SYSTEM: BUILDING CAPABILITIES AND CONFIGURATION OPTIONS

Abstract. Because it's very important to have a really educational parallel computer system for study architecture and programming techniques of computer system, which have a few process and Distributed Shared Memory. For example, the efficiency of this systems can be low. Possibilities of building educational cluster investigate in the work, based on the available Hardware and Software, in order to justify choice of the best solution, by the measure of efficiency/ cost of a computer system.

Keywords: cluster system, supercomputer, compute node, parallel programming.

Автори: Тиртышников А.И., Мильченко Т.О.

УЧЕБНАЯ ПАРАЛЛЕЛЬНАЯ ВЫЧИСЛИТЕЛЬНАЯ СИСТЕМА: ВОЗМОЖНОСТИ ПОСТРОЕНИЯ И ВАРИАНТЫ КОНФИГУРАЦИИ

Аннотация. Для изучения архитектуры и методов программирования компьютерных систем крайне желательно наличие реальных учебных параллельных вычислительных систем, имеющих несколько процессоров и распределенную память. При этом, например, производительность таких систем может быть невысокой. В работе исследуются возможности построения учебного кластера на основе доступного аппаратного и программного обеспечения с целью обоснования выбора оптимального по критерию эффективность / стоимость варианта учебной параллельной вычислительной системы.

Ключевые слова: кластерная система, суперкомпьютер, вычислительный узел, параллельное программирование