УДК 004.75:004.656(045)

Жуков І. А., д-р техн. наук
Іванкевич О. В., канд. техн. наук

АНАЛІЗ ВИКОРИСТАННЯ КЛАСТЕРНИХ ТЕХНОЛОГІЙ У СИСТЕМАХ КЕРУВАННЯ РОЗПОДІЛЕНИМИ БАЗАМИ ДАНИХ НА СУЧАСНИХ АВІАПІДПРИЄМСТВАХ

**Інститут комп'ютерних технологій
Національного авіаційного університету**

Проаналізовано сучасні кластерні системи, призначені для обробки великих обсягів даних. Досліджено особливості підвищення продуктивності інформаційних систем авіапідприємств за використання кластерних систем. Розглянуто мережеве обладнання та програмне забезпечення кластерів. Показано, що системи на базі кластерних технологій можуть ефективно обробляти великі масиви інформації для таких завдань, як моделювання траєкторії літаків, обробка даних чорних ящиків, обчислення великих масивів економічної інформації.

Вступ

При створенні сучасних інформаційних систем, що впроваджені на авіапідприємствах та під час своєї роботи використовують бази даних (БД) великого обсягу, для підвищення продуктивності додатків, забезпечення високої доступності даних, а також високої масштабованості систем використовуються кластерні технології [1−3].

Кластерні системи можуть бути використані для побудови недорогих, але потужних систем БД (СБД) [1, 2].

Кластери є альтернативою монокорпусним суперкомп'ютерам з оригінальною закритою архітектурою. Побудовані на базі компонентів, що серійно випускаються, вони широко застосовуються для здійснення високопродуктивних обчислень, забезпечення доступності й масштабованості.

Мета роботи

Метою статті є огляд кластерів, призначених для паралельної обробки великих обсягів даних та систем розподілу навантаження на вузлах кластера при обробці великих обсягів даних на підприємствах цивільної авіації. Досліджено особливості підвищення продуктивності кластерів за використання розподілених БД.

Кластер складається з незалежних комп'ютерів, зв'язаних між собою каналами передачі даних [2, 3]. Усі його підсистеми доступні у єдиному адміністративному домені, і керування ним виконується, як єдиною комп'ютерною системою. Вузли кластера - універсальні комп'ютери, які є серійними та здатні працювати самостійно. Вузли можуть бути одно- або багатопроцесорними. У класичній схемі всі вузли підчас роботи з додатками розділяють зовнішню пам'ять на масиві жорстких дисків, використовуючи внутрішні *HDD* для більш спеціальних функцій. Для міжвузлової взаємодії звичайно застосовується стандартна мережева технологія, проте не виключаються окремо розроблені канали зв'язку. Кластерна мережа є відокремленою – вона ізольована від зовнішнього мережевого середовища.

Аналіз існуючих технологій кластерів

Можна виділити три категорії кластерів за допомогою яких можна побудувати сучасні СБД, та які визначаються характером і призначенням додатків, з якими ці СБД будуть працювати [3, 4]. Кластери високої готовності (*High Availability, HA*), або відмовостійкі проектуються для забезпечення кінцевим користувачам безперебійного доступу до даних або до сервісів. Зазвичай, один екземпляр додатку працює на одному вузлі, а коли той стає

недоступним, керування додатком перехоплюється іншим вузлом (рис.1).

Така архітектура дозволяє проводити ремонт і профілактичні роботи, не зупиняючи сервіси. Якщо ж один вузол виходить з ладу, сервіс може бути відновлений без збитку для доступності інших. У цьому випадку продуктивність системи буде знижено.

Кластери високої готовності є якнайкращим вибором для забезпечення роботи критично важливих сервісів СБД, електронної пошти, файл-, прінт- і вебсерверів, а також серверів додатків. На відміну від розподілених і паралельних обчислень, ці кластери легко і прозоро включають додатки, що є у організацій, які не орієнтовані на кластери. Це дозволяє розширювати мережу у міру зростання системи.

Кластери такого типу використовують в авіакомпаніях та на підприємствах, що контролюють повітряний простір країни, зокрема в авіакомпанії «Міжнародні авіалінії України» (МАУ), яка має 17 сучасних літаків Боїнг-737 та виконує рейси у більш ніж 70 країн світу [5]. Авіакомпанія використовує кластер типу *HA*, зібраний на основі вузлів, кожен з яких є двопроцесорною системою на базі *AMD Athlon MP* 1500+ для обробки даних чорних ящиків всіх літаків після кожного рейсу. На основі обробленої інформації діагностують проблеми, що можуть виникнути в обладнанні літака та контролюють майстерність пілотів. Уся система складається з чотирьох вузлів і одного керуючого комп'ютера. Вузли сполучені між собою комутатором *Gigabit Ethernet*. Сам кластер розміщений в двох 45-юнітових шафах з вмонтованими блоками вентиляторів і термовимикачами. Вузли і координувальний комп'ютер розміщені в корпусах *4u Rack Mount* на телескопічних кріпленнях. Для візуалізації роботи вузлів кластера використовується восьмипортовий перемикач монітора/клавіатури/миші *ATEN Master View Cs*-9138.

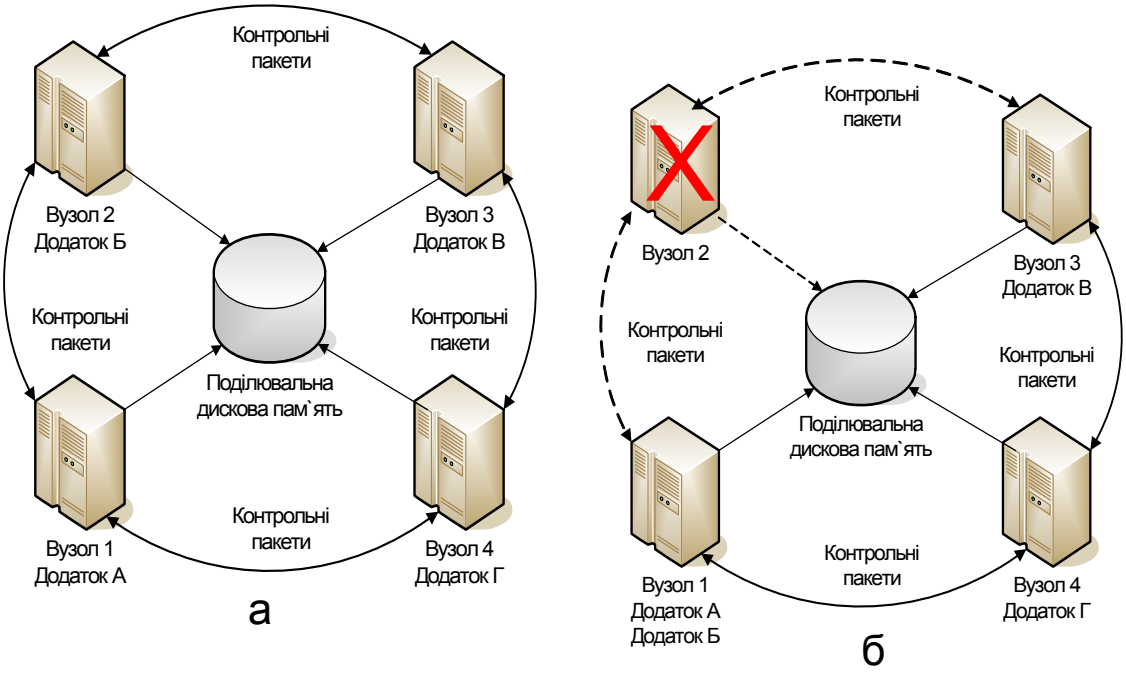

Рис. 1. Відмовостійкий кластер:
а – робота у штатному режимі;
б – робота після відмови

Кластери балансування навантаження (*Load Balancing, LB*) розподіляють вхідні запити між багатьма вузлами, на яких працюють однакові програми або розміщений один і той самий контент (рис. 2).

Кожен вузол може обробляти запити до одного і того ж додатку або контенту. Якщо будь-який із вузлів виходить з ладу, запити перерозподіляються між тими, що лишилися.

У типовому випадку такі кластери використовуються для обробки великих обсягів даних та веб-хостінгу. Розглянуті кластерні технології можна об'єднувати для збільшення надійності, доступності і масштабованості додатків.

Такі кластери використовуються в авіакомпаніях та міжнародних аеропортах, зокрема системи *HPC* використовують у міжнародному аеропорту Борисполь [6] та Франкфуртському міжнародному аеропорту [7]. На цих авіапідприємствах кластерні технології використовують під час оперативного контролю руху великої кількості літаків. У міжнародному аеропорту м. Франкфурт використовують обчислювальний кластер *LB*, робота якого забезпечується 69 серверами *IBM pseries*, 20 *IBM*-нодамі *SP* під керуванням *ОС AIX,* 15 *SUN (ОС Solaris)*, 100 лезовими *intel*-системами (*Linux/windows*) також від *IBM* і 90 *HP Proliant (Windows)*, розміщеними в 160 стійках. Система зберігання (без врахування пристроїв, призначених для резервного копіювання) має загальний обсяг близько 50 ТБ, чотири дискові масиви *IBM ESS* 800 ємністю 21,5 ТБ, 20-терабайтовий масив *IBM Ds*4800 і 6-терабайтовий *HP*.

Рис. 2. Кластер балансування навантаження

Кластери для високопродуктивних обчислень (*High-Performance Cluster, HPC*). Традиційно паралельні обчислення виконувалися на мультипроцесорних системах, спеціально спроектованих для цього. В таких системах безліч процесорів розділяли загальну пам'ять і шинний інтерфейс в межах одного комп'ютера. З появою високошвидкісної комутаційної технології стало можливим об'єднувати комп'ютери в кластери для паралельних обчислень.

Паралельний кластер використовує велику кількість вузлів для розпаралелю-

вання обчислень під час розв'язку специфічного завдання. На відміну від кластерів, що використовують балансування навантаження та кластерів високої готовності, які розподіляють запити та завдання між вузлами, що обробляють їх у цілому, у паралельнім середовищі запит підрозділяється на безліч підзавдань, а ті, у свою чергу, розподіляються для обробки між вузлами усередині кластера.

Паралельні кластери застосовуються переважно для додатків, що вимагають інтенсивних математичних обчислень. У цивільної авіації це моделювання траєкторії літаків, обробка даних чорних скринь, обчислення великих масивів економічної інформації тощо.

Базові компоненти кластерів розбиваються на кілька категорій: безпосередньо вузли, кластерне програмне забезпечення (ПЗ), виділена мережа, яка робить обмін даними між вузлами відповідними мережевими протоколами

Нині замість традиційних серверних корпусів використовують вмонтовані в одну стійку мультипроцесорні системи й лезові сервери, які забезпечують вищу процесорну щільність в умовах дефіциту простору. Продуктивність процесорів, пам'яті, швидкість доступу до жорстких дисків і їх ємність значно збільшилися, однак при такому експонентному рості в деяких випадках швидкодії вартість цих технологій суттєво знизилася.

У типовому випадку вузол у кластері може бути керуючим (головним) або обчислювальним (підлеглим) (рис. 3). Головний вузол може бути тільки один. Він відповідає за роботу кластера та є ключовим для кластерного ПЗ проміжного шару, процесів маршрутизації, диспетчеризації та моніторингу стану кожного обчислювального вузла. Останні виконують обчислення й операції із системою зберігання даних та є повнофункціональними автономними комп'ютерами.

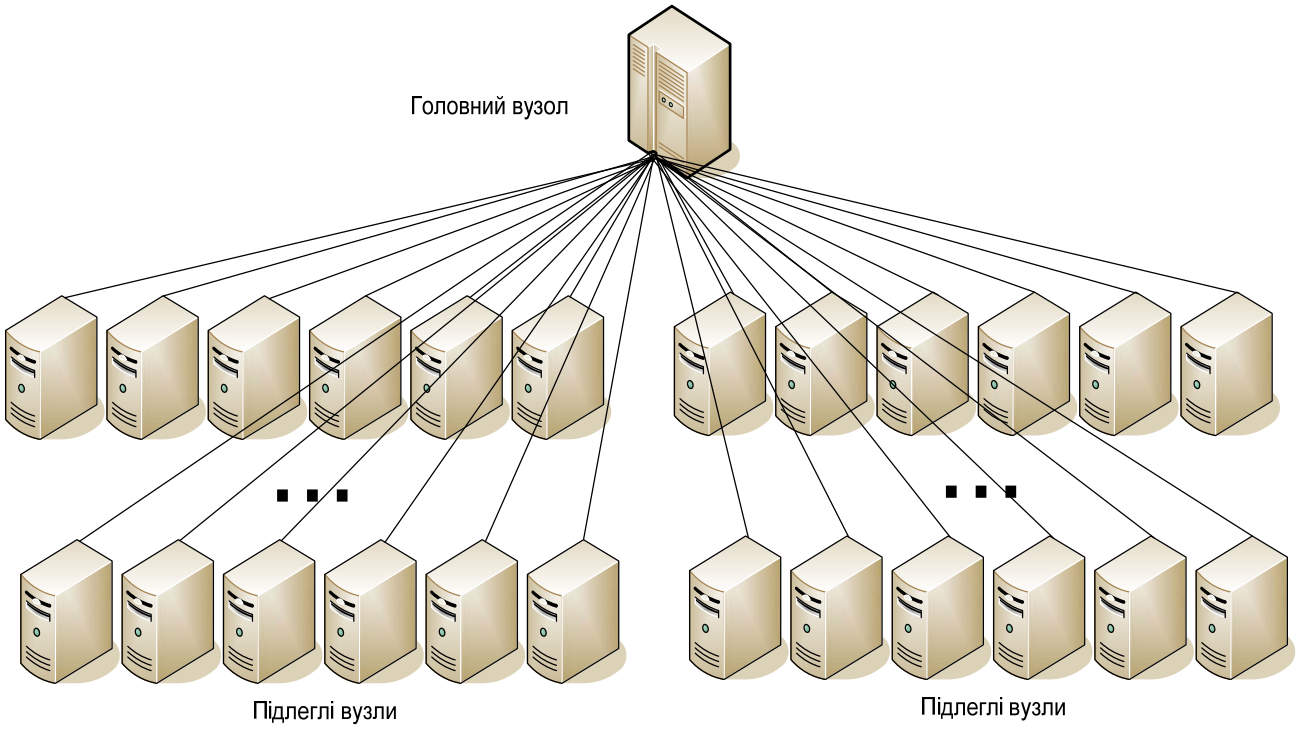

Рис. 3. Кластер для високопродуктивних обчислень

Кластери такого типу використовуються передусім у галузі літакобудування та проектування авіаційних систем. Таку систему використовують на НВО «Салют», що розробляє газотурбінні двигуни для літаків [8]. Обчислювальний кластер

цього підприємства будується за класичною схемою і складається з 50 серверів *Rx200 S3* виробництва *Fujitsu-siemens Computers* і комутаційного устаткування *Cisco Catalyst*. Кластер використовує нові чотириядерні процесори *Intel Xeon X5355* з частотою 3ГГц [8].

Вживання кластерних обчислювальних технологій у процесі створення авіаційних і газотурбінних двигунів дозволяє збільшити конкурентоспроможність виробу шляхом істотного скорочення часу розробки та економії матеріальних ресурсів. Кількість дослідних зразків виробів і натурних випробувань зводиться до мінімуму. Сумарна економія становить мільйони доларів і тисячі людино-годин.

Програмне забезпечення

Як і у звичайному комп'ютері, операційна система (ОС) кластера є серцем кожного його вузла. Вона присутня при будь-якій дії користувача: звертанні до файлової системи, відправленні повідомлень або старту додаткового процесу. Користувачі можуть вибирати різні парадигми програмування або ПЗ проміжного шару, але кластерна ОС спільна для всіх.

Основна роль кластерної ОС полягає в першу чергу в тому, щоб мультиплексувати безліч користувацьких процесів на єдиний набір апаратних компонентів (керування ресурсами) і забезпечити придатні абстракції для високорівневого ПЗ. Деякі з цих абстракцій включають захист границь пам'яті, координацію процесів, потоків і комунікацій та керування пристроями. Більшість специфічних для кластера функцій виконується ПЗ проміжного шару. Операційна система кластера досить складна, і не завжди зрозуміло, як зроблені зміни вплинуть на інші системи. Тому необхідні модифікації краще проводити на рівні ПЗ проміжного шару, причому додана в нього нова функціональність може бути перенесена на інші ОС.

Адміністратори й користувачі бачать кластер як єдину обчислювальну систему. Це досягається за допомогою образу єдиної системи (*Single System Image, SSI*). Саме кластер приховує неоднорідну й розподілену природу наявних ресурсів і використовується користувачами і додатками як єдиний обчислювальний ресурс. *SSI* може бути реалізовано на одному або декількох з таких рівнів: апаратному, ОС, ПЗ проміжного шару або додатка.

Системи *Digital/Compaq Memory Channel* і *Distributed Shared Memory* забезпечують *SSI* на апаратному рівні й дозволяють користувачам бачити кластер як систему з поділюваною пам'яттю. ОС *SCO Unixware Nonstop Cluster, Sun Solaris-mc, GLUNIX і MOSIX* підтримують *SSI* на рівні ядра.

Мережеве обладнання кластерів

Створення загальнодоступних кластерів стало можливим тільки завдяки адекватним мережевим технологіям для міжвузлових комунікацій. Загальнодоступні кластери включають одну або більше виділених мереж для передачі пакетів повідомлень всередині розподіленої системи.. Сьогодні у розробників кластерів є широкі можливості для вибору мережевої технології. Практика дає приклади побудови досить ефективних кластерів з використанням недорогого мережевого обладнання, яке можна побачити у звичайній локальній обчислювальній мережі. У той же час окремі мережеві продукти, спеціально розроблені для кластерних комунікацій, можна порівняти по вартості з робочими станціями. Вибір мережевої технології залежить від ряду факторів: ціни, продуктивності, сумісності з іншим кластерним обладнанням та ПЗ, а також від комунікаційних характеристик додатків, які будуть виконуватися на кластері.

Продуктивність мережі в загальному випадку описується в термінах латентності та смуги пропускання. Латентність – відрізок часу від запиту даних до їх одержання, або час, за який вони передаються від одного комп'ютера іншому, включаючи непродуктивні витрати ПЗ на формування повідомлення та час передачі бітів. В ідеалі в додатках, написаних для кластерів, обмін повідомленнями має бути мінімальним. Якщо програма посилає

велику кількість коротких повідомлень, тоді його продуктивність буде залежати від латентності мережі. Якщо ж відбувається обмін довгими повідомленнями, то основний вплив на цей параметр надасть їй пропускну здатність на високому рівні. Очевидно, продуктивність програми буде найкращою при низькій латентності та широкій смузі пропускання. Для задоволення цих двох вимог необхідні ефективні комунікаційні протоколи, що мінімізують обсяг службових даних, а також швидкі мережеві пристрої. Комунікаційні, або мережеві, протоколи визначають правила й угоди, які будуть використовувати два або більше комп'ютерів в мережі для обміну інформацією. Вони можуть бути з встановленням або без встановлення з'єднання, надавати різний рівень надійності - з повною гарантією доставки в порядку слідування пакетів і без такої, синхронні (без буферізації) і асинхронні (з буферізацією). Для кластерних комунікацій застосовуються як традиційні мережеві протоколи, розроблені спочатку для Інтернету (IP), так і створені спеціально.

Крім цього, є два відносно нових стандарти, які також спеціально призначені для кластерів. До протоколів *IP*, які є специфічним для кластерів, належать *Active Messages, Fast Messages, Virtual Memory-Mapped Communication system, U-Net ma Basic Interface for Parallelism*.

Для кластерних комунікацій використовується стандарт *Virtual Interface Architecture (VIA)* та стандарт для поділюваних підсистем зберігання *InfiniBand. VIA* – це комунікаційний стандарт, що поєднує кращі досягнення різних проектів, був створений консорціумом академічних та індустріальних партнерів, що включає *Intel, Compaq і Microsoft. VIA* базується на концепції віртуального мережевого інтерфейсу. Стандарт передбачає, що перед відправленням повідомлення передавач і приймач буфера повинні бути виділені і прив'язані до фізичної пам'яті. Після того як буфер, та пов'язані з ним структури даних сформовані, ніяких системних викликів не потрібно. Операції прийому і відправки в додатку користувача складаються з запису дескриптора в чергу. Додаток може вибирати, чекати підтвердження завершення операції або продовжувати основну роботу, поки повідомлення обробляється.

Хоча *VIA* може бути доступний прямо для прикладного програмування, багато хто з розробників систем вважають, що це занадто низький рівень для додатків, так як останні повинні бути відповідальними за розподіл частини фізичної пам'яті і стежити за її ефективним використанням. Передбачається, що більшість виробників ОС, ПЗ проміжного шару забезпечать інтерфейс з *VIA*, який буде підтримувати прикладне програмування. Так, зараз більшість постачальників баз даних пропонують версії своїх продуктів, що працюють поверх *VIA*. Швидко стає доступним і інше кластерне ПЗ, наприклад, файлові системи.

Стандарт *InfiniBand* використовується фірмами *Compaq, Dell, HP, IBM, Intel, Microsoft і Sun Microsystems*. Архітектура *InfiniBand* замінює розподільну шину, що є стандартом для високошвидкісної послідовної системи вводу-виводу в сучасних комп'ютерах, базуватиметься на механізмі каналів комутаційної фабрики. Усі системи та пристрої підключають до фабрики за допомогою канального адаптера хоста (*Host Channel Adaptor, HCA*), який забезпечує з'єднання центрального процесора хоста зі структурою *InfiniBand*, або канального адаптера цільового вузла (*Target Channel Adaptor, TCA*), з'єднуючого *InfiniBand* з іншими пристроями вводу-виводу типу *Ethernet, Fibre Channel* або з системами зберігання даних. Канал *InfiniBand* дуплексний і працює з пропускною здатністю 2,5 Гб/с в одному напрямку в топології «точка-точка». До того ж *InfiniBand* підтримує віддалений прямий доступ до пам'яті, який дозволяє одному процесору читати або писати в пам'ять іншого.

Мережеве обладнання, що підтримує міжвузловий обмін може оцінюватися за допомогою чотирьох категорій залежно

від того, чи виконується підключення до шини вводу-виводу або до шини пам'яті, і від основного методу комунікацій – за допомогою повідомлень або вашої дискової пам'яті.

Із чотирьох категорій взаємоз'єднань найпоширенішими є системи на базі повідомлень і з підключенням до шини вводу-виводу, оскільки в цьому випадку інтерфейс з комп'ютером найбільш зрозумілий. Шина вводу-виводу має апаратне переривання, яке інформує процесор, про готовність даних для читання. Такі системи реалізовані у всіх широкодоступних мережевих технологіях, а також у низці останніх продуктів, розроблених спеціально для кластерних обчислень.

У системи з підключенням до шини вводу-виводу і з вашої дискової пам'яттю входять комп'ютери з вашої дискової підсистеми. Приєднання до пам'яті менш поширене, оскільки шина пам'яті має індивідуальний дизайн для кожного типу комп'ютерів. Однак багато таких систем реалізуються з допомогою ПЗ або за допомогою механізму відображення портів вводу-виводу в пам'ять, як *Memory Channel.*

Відомі гібридні системи, які комбінують особливості декількох категорій, наприклад, InfiniBand дозволяє посилати як дані на диск, так і повідомлення інших вузлів. Аналогічно *Scalable Coherent Interface (SCI)* може також використовувати обидва механізми обміну.

Висновки

Існує широкий спектр кластерів, які можна використовувати під час виконання великих обсягів обчислень. Вони відрізняються типом і швидкодією процесорів, розміром поділюваної вузлами пам'яті, технологією взаємозв'язку вузлів, моделями й інтерфейсами програмування. Однак результат, що досягається з їхньою допомогою, значно залежить від особливостей додатків, які планується на них розгорнути. Існує можливість побудови на базі кластерних технологій СБД, які зможуть ефективно обробляти великі масиви інформації таких задач, як моделювання траєкторії літаків, обробка даних чорних ящиків, обчислення великих масивів економічної інформації тощо.

Список літератури

1. *Жуков И.А., Гуменюк А.В.* Перспективы использования кластерных вычислительных систем в авиации // Вісник Київського міжнародного університету цивільної авіації. – К.: КМУЦА, 1999.– № 2. – С. 96–100.

2. *Креденцар С.М.* Перспективы применения параллельных вычислений и кластерных вычислительных систем в системах отображения воздушной обстановки // Матеріали VII Міжнародної науково-технічної конференції "АВІА-2006", 25-27 вересня 2006 р.– К.: НАУ, 2006. – Т. 1. – С. 21.113–21.116.

3. *Корочкин А.В.* Организация вычислений в кластерных системах с многоядерной архитектурою // Проблеми інформатизації та управління: збірник наукових праць. – К.: НАУ, 2008. – Вип. 1 (23). – С. 143–145.

4. *Гуменюк В.О.* Технології кластерних архітектур // Проблеми інформатизації та управління: Збірник наукових праць. – К.: НАУ, 2004. – Вип. 10. – С. 151–156.

5. Авіакомпанія «Міжнародні авіалінії України» // *http://www.flyuia.com/ua/main.html.*

6. Державне підприємство «Міжнародний аеропорт «Бориспіль» // *http://www.airport-borispol.kiev.ua/.*

7. *Кученко Ю.* Бесперебойная ИТ-инфраструктура Франкфуртского аэропорта // Компьютерное обозрение. – К., 2008. – Вип. 19 (636). – С. 14–17.

8. Науково-виробниче об'єднання «Салют» // *http://www.npo-saturn.ru/!new/?act=gm_look&id=11879540.*