УДК 004.32

О. Я. Матов, І. О. Храмова
Інститут проблем реєстрації інформації НАН України
вул. М. Шпака, 2, 03113 Київ, Україна

Проблеми використання і математичне моделювання хмарних обчислень для інтегрованої інформаційно-аналітичної системи державного управління

Розглянуто необхідність і можливість застосування глобальних інформаційних технологій для інтеграції відомчих систем в інформаційно-аналітичній системі державного управління. Запропоновано аналітичні моделі для оцінки обчислювальних процесів вузлів хмарних обчислень.

Ключові слова: *архітектура хмарних обчислень, інтеграція, моделі обслуговування.*

Вступ

Вимоги до обчислювальних інфраструктур щодалі зростають із поширенням глобальних державних програм на зразок концепцій електронного урядування (e-Government), електронних бібліотек (e-Library), електронної науки (e-Science) та ін., що вже набули поширення в світі та в Україні.

Колишні способи надання, виробництва та споживання інформації, що були перенесені спочатку до відомчих інформаційно-аналітичних систем, а надалі і в Інтернет-середовище майже без змін, більше не здатні забезпечувати її ефективне використання. Практика свідчить, що органи державного управління витрачають значні кошти на розробку та придбання технологічних рішень, не отримуючи очікуваної віддачі від їхнього впровадження.

У той час, як хтось уже просунувся в своєму розумінні того, як проектувати і діяти, щоб повторно використовувати та застосовувати вже наявні рішення, більшість продовжує вкладати свої ресурси в пошуки відповідних засобів і інноваційних методів. Фундаментом будь-якої інформаційно-аналітичної системи є комплекс потужних баз даних (БД) з налаштованою на проблемне середовище структурою. Такі комплекси, якщо вони реалізовані як сховища даних, у сутності, — це сімейства БД, що містять взаємозалежну інформацію. Важливим кроком уперед від «різноцвіття» представницьких вебсайтів до певного впорядкування накопичених держустановами терабайтів електронних документів шляхом їхньої каталогізації було створення мережі порталів урядових установ. Проте, для вирішення глобальних завдань електронного урядування замало тільки забезпечувати споживача інформацією.

Архітектура спільного користування урядовими інформаційними ресурсами

Нові проекти інформатизації органів державного управління часто створюються з

попереднім аналізом або оцінкою вже існуючих програмних продуктів. Найчастіше ці аналітичні огляди, на жаль, обмежуються винятково галузевим доменом і апелюють до однакових проблем. Як результат, у державних установах існує велика кількість окремих процесів, що дублюються, а також технологічно еквівалентних функціональних систем, що, як правило, вирішують однакові з погляду інформаційних технологій завдання, обробляють ті самі набори даних і збирають однакову за контекстом інформацію (прикладом може правити славнозвісний Єдиний державний реєстр підприємств і організацій України, Класифікатор об'єктів адміністративно-територіального устрою України, електронні карти з інформаційними шарами до них, збірки державних нормативних документів і багато інших), які, тим не менше, є несумісними, коли постає питання про інтеграцію інформаційних ресурсів для реалізації горизонтальних міжвідомчих інформаційних процесів. У підсумку: час триває, один за одним з'являються інші проекти інформаційних та інформаційно-аналітичних систем, цілі яких перетинаються, а плани реалізації перебувають у жорсткій конкуренції за ресурси. Все це не стільки допомагає вирішувати проблеми суспільства і держави, скільки додає зайвих труднощів.

Рамкова архітектура (електронного) суб'єкта державного управління мала б забезпечити структуру (каркас) для інформаційних систем у масштабах уряду України, що дозволило б міністерствам і відомствам розділяти (у значенні спільно використовувати) загальні дані, інформацію та взаємопов'язані ділові процеси всіх суб'єктів державного управління.

Іншим ключовим фактором, що також відіграє роль двигуна зростання ефективності одержуваних результатів діяльності в державному управлінні, є впровадження і поширення в практику державного управління методів добування даних (data mining) та інтелектуального аналізу даних, а також відповідних інструментальних засобів, що є досить витратним заходом. Тому вимога спільного використання подібних, достатньо коштовних, засобів і напрацьованих ділових процесів у розподіленій мережній структурі є цілком природною.

Рішенням може бути створення нового абстрактного рівня архітектури державних і відомчих інформаційних ресурсів, що надаватимуться для спільного користування, та реєстру урядових і відомчих інформаційно-аналітичних ресурсів, які відповідатимуть цьому рівню (рис. 1). Новий архітектурний рівень мав би охопити питання міжмережної взаємодії (interconnectivity), інтеграції даних, інформаційного доступу і управління контентом. Згадані питання є необхідними для того, щоб підтримувати урядовий рівень виконання трансакцій та надання інформаційно-комунікаційних послуг, а також для того, щоб здійснити інтеграцію інформаційних систем у межах усіх урядових відомств. Оскільки такий загальнодержавний інформаційний ресурс має служити багатьом державним установам, необхідно піклуватися й про те, щоб він мав високий рівень доступності та ніколи не виходив з ладу. Це можливо, якщо в основу рішення покласти сучасні глобальні архітектури, які дозволяють організацію розподілених обчислень із використанням обчислювальних мереж (метакомп'ютинг).

На сьогодні мережі вже довели свою практичну корисність як засіб глобальної доставки різних форм інформації. Проте потенціал застосування мереж значно ширший: вони мають розглядатися ще й як засіб організації обчислень наступної генерації.

Технології, що підтримують спільне й скоординоване використання різних ресурсів у динамічних розподілених віртуальних організаційних структурах, отримали назву

«хмарні обчислення» (cloud computing)[1] і ставлять за мету створення з географічно і організаційно розподілених компонентів віртуальних обчислювальних систем, що достатньо інтегровані, щоб надати бажану якість обслуговування. Термін «хмарні обчислення» з'явився на початку 1990-х рр. як метафора, заснована на зображенні Інтернету на діаграмі комп'ютерної мережі у вигляді хмарки, із певною асоціацією про таку ж легкість доступу до обчислювальних ресурсів.

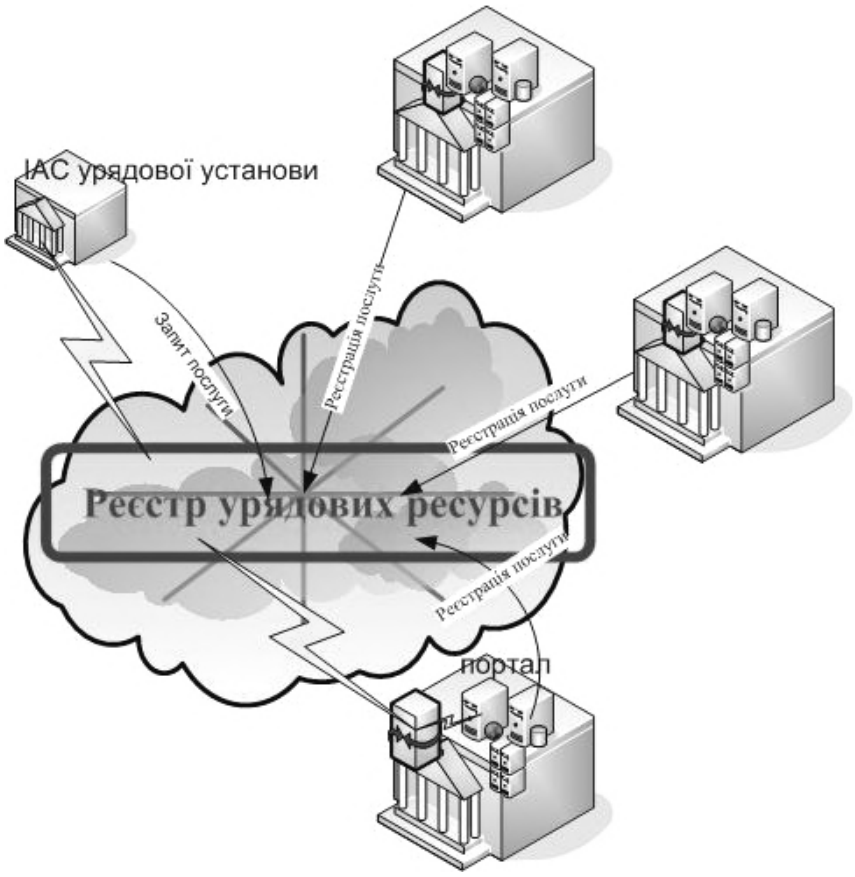

Рис. 1. Архітектура спільного користування урядовими інформаційними ресурсами

Хмарні обчислення (ХО), згідно з документом, виданим IEEE в 2008 р. [1], — це «парадигма, в рамках якої інформація постійно зберігається на серверах в Інтернеті й тимчасово кешується на клієнтській стороні, наприклад, на персональних комп'ютерах, ігрових приставках, ноутбуках, смартфонах і т.ін.». Обчислювальна інфраструктура об'єднує безліч ресурсів різних типів (процесори, програмне забезпечення, довготривалу й оперативну пам'ять, сховища і бази даних, мережі), доступ до яких користувач може отримати з будь-якого місця, незалежно від його розташування. Ідея хмарних обчислень, як наступник GRID-технологій інтеграції ресурсів [2], виникла у той же час, на який припали поширення персональних комп'ютерів, розвиток Інтернету і техноло-

[1] Хоча це поняття вже є сталою словосполукою, інколи «cloud computing» перекладається з англійської також поняттями « хмарна обробка даних» та «розсіяні обчислення». Останній термін вбачається найбільш відповідним сутності концепції.

гій пакетної передачі даних на основі оптичного волокна (SONET, SDH і ATM), а також технологій локальних мереж (Gigabit Ethernet). Смуга пропускання сучасних комунікаційних засобів стала достатньою, щоб за необхідності (за попередньою домовленістю) залучити ресурси інших комп'ютерів і, навіть, комп'ютерних ресурсів інших власників. Користувач має доступ до власних даних, але не може управляти і не повинен піклуватися про інфраструктуру, операційну систему і власне програмне забезпечення (ПЗ), з якими він працює. Переваги хмарних обчислень досить переконливі: широкі можливості віртуалізації ресурсів за необхідності, висока доступність, легше адміністрування програмних активів, «еластичне» масштабування. Сьогодні хмарні обчислення називають перспективним трендом у розвитку інформаційних технологій.

Із поняттям хмарних обчислень пов'язують такі інформаційні технології надання послуг (або сервісів), як «Програмне забезпечення як сервіс» («Software as a Service» або «SaaS»), «Інфраструктура як сервіс» («Infrastructure as a Service» або «IaaS»), «Платформа як сервіс» («Platform as a Service» або «PaaS») та ін.

SaaS — модель розгортання застосування, яка передбачає надання застосування кінцевому користувачеві подібно до того, як надається послуга на вимогу (Service on Demand). Доступ до такого застосування здійснюється за допомогою мережі, а найчастіше за допомогою Інтернет-браузера.

IaaS — модель надання комп'ютерної інфраструктури як сервісу. Замість закупівлі серверів, ПЗ, спеціального мережевого устаткування, користувач може отримати ці ресурси у вигляді аутсорсингових (outsource) послуг.

PaaS — модель надання обчислювальної платформи як сервісу в мережі, який пропонує розгортання та підтримку веб-сервера застосувань і сервісів розробки без необхідності покупки й управління програмним або апаратним забезпеченням.

Ці технології при спільному використанні дозволяють користувачам хмарних обчислень скористатися обчислювальними потужностями, ПЗ і сховищами даних, які за допомогою певних технологій віртуалізації та високого рівня абстракції надаються їм як послуга.

Для управління елементами інфраструктури використовується спеціалізоване ПЗ середнього або проміжного рівня, що узагальнено називають «middleware control». Це ПЗ надає ключові сервіси, такі як узгодженість, транзакційність, багатопоточність і обмін повідомленнями для застосувань, побудованих на основі сервісно-орієнтованої архітектури (SOA). У функції middleware control також входять сервіси безпеки і високої доступності.

Відрізняють два види хмарних обчислень:

— публічні хмари (загальнодоступні) — абонентами можуть бути будь-які організації і індивідуальні користувачі у великій кількості. Інформаційні технології (бізнес-системи, веб-сайти) надаються на базі «комунальних» (multi-tenancy) інфраструктур з великими можливостями масштабування, які в інших рішеннях були б недоступні. Як приклади можна навести онлайнові сервіси Amazon Elastic Compute Cloud (Amazon EC2) [4] і Amazon Simple Storage Service (Amazon S3) [5], сервіси Google Apps — розширені можливості таких всесвітньо відомих продуктів Google, як Gmail, Документи, Сайти [6] і багато інших.

— приватні хмари — абонентами є корпоративні користувачі (організації і їхні підрозділи, об'єднані загальною діяльністю), а процеси та дані не виходять за межі корпоративної мережі.

Приватна хмара є більш безпечною IT-інфраструктурою, захищеною файрволом, яка контролюється і експлуатується на користь однієї спільноти корпоративної спів-

праці. Приватні хмарні обчислення поєднують властивості керованості та безпеки з гнучкістю, необхідною для ділових (управлінських) нововведень. До того ж, вони значно скорочують витрати замовника. Приватні хмари вирішують ряд серйозних проблем, від яких не застраховані публічні хмарні обчислення, а саме: безпека, конфіденційність даних, час очікування, дотримання вимог державних і галузевих регуляторів.

Крім згаданих, можна визначити ще один, гібридний, вид ХО — «віртуальна приватна хмара», тут мається на увазі те, що провайдер використовує публічну хмарну інфраструктуру для організації інфраструктури приватної хмари. При такій організації, дані клієнта частково зберігаються і обробляються за рахунок ресурсів власної інфраструктури, а частина даних — за рахунок ресурсів зовнішнього провайдера. За приклад може правити веб-сервіс під назвою Amazon Virtual Private Cloud (Amazon VPC) компанії Amazon [7].

GRID як технологія створення розподіленої обчислювальної інфраструктури хмарних обчислень і засіб групового використання ресурсів

Останніми роками GRID, як технологія розподілених обчислень, зайняла власне місце і, розробивши достатньо розвинутий власний стандарт OGSA (Open Grid Services Architecture) [8], зараз уже розглядається як могутній еволюційний стрибок більшості з відомих нині розподілених технологій таких як Web, однорангові мережі, кластери і розподілені обчислення, технології віртуалізації.

Однак, крім загальних з іншими технологіями розподілених обчислень властивостей, GRID-технології мають відмінності [3], які роблять цю технологію дуже привабливою для багатьох секторів у галузі корпоративних обчислень [13–20]. Достатньо згадати, що серед відомих комерційних компаній, які мають відповідне ПЗ і підтримують стандарт OGSA, присутні Hewlett-Packard, IBM, Microsoft, Oracle, Sun та багато інших.

Спроби подолати обмеження поодинокої обчислювальної системи набули великої популярності. Доказом є те, що переважно всі європейські країни, Україна, Росія, Китай вже декілька років мають програми досліджень і проекти застосування GRID [9].

У тій частині, де GRID-технології збігаються із технологіями віртуалізації, і лежить площина організації хмарних обчислень.

Архітектура системи ХО, як інтегрованої інформаційно-аналітичної системи державного управління, концептуально надається сукупністю 3-х шарів. Основу всього складає «хмара обчислень і даних», тобто апаратні засоби обчислювальних машин і мереж передачі даних, на яких, власне, й виконується завдання користувача. Над цим розміщується «інформаційна хмара», тобто ті інформаційні бази даних (БД), до яких звертатимуться апаратні засоби та системи для операцій над даними. Завершує все «хмара знань», де високорівневі додатки мають вишукувати дані для отримання знань, що надалі ляжуть в основу процесів семантичного порозуміння та інтелектуального прийняття рішень. У загальному вигляді основні компоненти інфраструктури ХО наведено на рис. 2.

Запорукою ХО є програмне забезпечення, що здатне організувати та інтегрувати роздрібнені несумісні обчислювальні потужності в єдине ціле, автоматично підтримуючи міжмашинні зв’язки, потрібні для створення з різнорідних обчислювальних та інформаційних ресурсів прозорого індивідуального «механізму», необхідного і достатнього для виконання нагального завдання, й надання користувачеві або порталу безпечного доступу до цього «механізму». Центральним елементом у взаємодії трьох ключо-

вих взаємодіючих сутностей ХО-систем — користувачів, даних і ресурсів — є систематизовані метадані, що описують ці сутності, та роблять можливою автоматичну їхню взаємодію. Метадані надають індивідуальні «агенти» кожної із сутностей, у той час, як процесами міжмашинної взаємодії та підтримки домовленостей між окремими сутностями керують «брокери». Мережні «агенти» відповідають за оптимізацію маршрутизації та підтримку рівня очікуваної якості послуг.

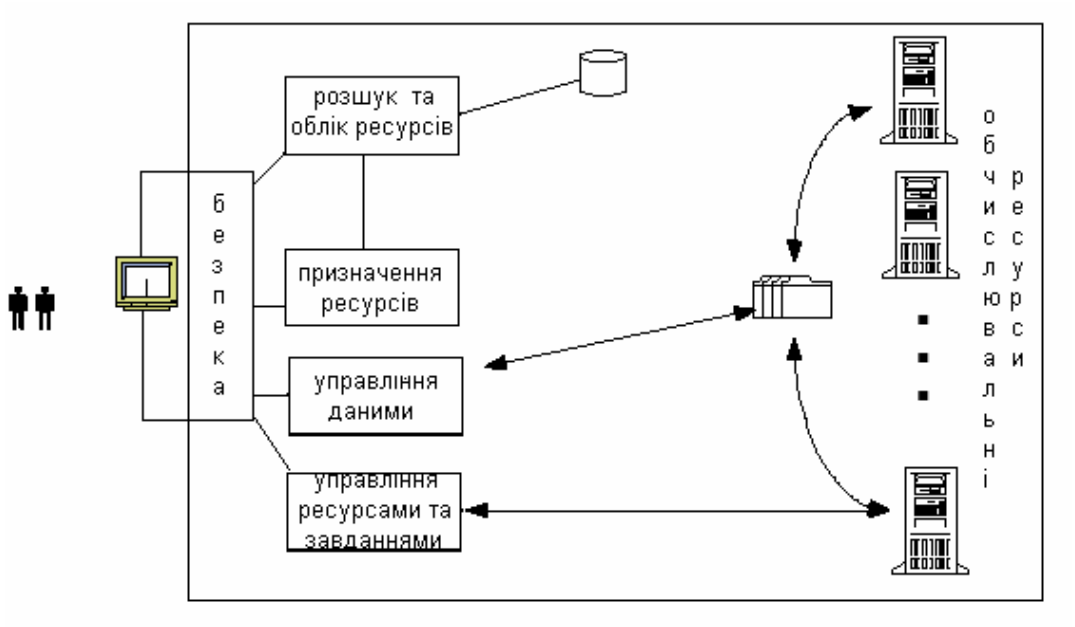

Рис. 2. Компоненти ХО-інфраструктури

ХО, як ІТ, поділяє концепцію GRID, що будь-який її компонент (обчислювальний ресурс, ресурс зберігання інформації, мережа, додаток, база даних і т.ін.) має розглядатися як сервіс (послуга), тобто, доступна мережею сутність (об'єкт), що забезпечує певні можливості шляхом обміну повідомленнями. Прийняття сервіс-орієнтованої моделі за базисну дозволяє віртуалізувати будь-який компонент обчислювального оточення. Основу ХО-сервіса, як і GRID-сервіса складають одні й ті ж самі 4 «кити» [10]: XML (eXtensible Markup Language), який набув популярності не тільки як мова розмічування для обміну даних, але і як формат для напівструктурованих даних; SOAP (Simple Object Access Protocol) — протокол, незалежний від протоколу нижчого рівня, який забезпечує засоби передачі повідомлень між постачальником і споживачем служби; WSDL (Web Services Description Language) специфікація мови опису веб-служби; WSIL (Web Services Inspection Language) специфікація мови для визначення місцезнаходження описів веб-служб, що постачаються [10].

Визначення ефективності організації хмарних обчислень

Кількісно спосіб організації обчислювального процесу (ОП) може бути оцінений показниками ефективності. Багатогранність проблеми визначення ефективності ОП, складність обліку великої кількості факторів, обумовлюють наявність широкого діапазону показників ефективності способів організації ОП.

Найчастіше показники ефективності поділяють на дві великі групи: показники, що базуються на оцінці середнього чи максимального часу перебування (затримки щодо

припустимих термінів) завдань у системі [11] та показники, що базуються на оцінці продуктивності структурно-функціональних компонентів. Останні характеризується різними факторами: кількістю та завантаженням задіяних ресурсів, часом їхніх простоїв, частотою конфліктів при звертанні до загальних обчислювальних ресурсів [11, 12] тощо.

Розглянемо першу групу показників, як таку, що наочно визначає ефективність обчислень.

Реальні системи, що виступають в нашому випадку сервісом вузла ХО, звичайно ж проектуються з відповідністю до підвищених вимог до надійності функціонування. Пристрої, що відмовили, як правило, відновлюються. Проте, поява відмов фізичних пристроїв призводить до погіршення характеристик функціонування вузла. Правила поведінки вузла по відношенню до завдання, коли відмова мала місце під час його обслуговування, сильно залежить від типу відмови. Наприклад, якщо відмовив процесор, завдання повинне бути виконано повторно, якщо під час завантаження/вивантаження даних — обслуговування може бути подовженим після відновлення. Проте, обмеження часу перебування завдання у системі, що накладаються класом систем, усе ж можуть привести до стану, який можна визначити, як відмова вузла. Варіанти резервування, що можуть бути вжиті, правила обслуговування за наявності відмов мають бути співставленні ще на етапі проектування. Вирішується таке завдання моделюванням вузла, з метою виявлення найбільш ефективного варіанта організації ОП, який забезпечив би найменше зниження продуктивності за наявності відмов.

У [11, 12] та деяких інших роботах пропонується оцінювати ефективність організації обчислювальних процесів за допомогою наступних функціоналів:

$$C^{(S)} = \sum_{i=1}^{n} \alpha_i \lambda_i v_i^{(S)}. \tag{1}$$

$$C'^{(S)} = \sum_{i=1}^{n} \alpha_i' \lambda_i P_i^{(S)}(> D_i), \tag{2}$$

де α_i, α_i' — відповідно штрафи за одиницю часу перебування завдання *i*-го типу в системі і його втраті внаслідок перевищення припустимого директивного часу D_i; λ_i — інтенсивність *i*-го потоку завдань; $v_i^{(S)}$ — середній час перебування завдань *i*-го типу в системі; $P_i^{(S)}(> D_i)$ — імовірність чекання завдання *i*-го типу понад припустимий директивний час D_i; n — кількість типів завдань; S — параметр, що характеризує варіант або спосіб організації обчислювального процесу; $C^{(S)}, C'^{(S)}$ — відповідно середні сумарні штрафи за перебування завдань у системі та їхньої втрати внаслідок перевищення припустимого директивного часу.

Показник ефективності (1) базується на припущенні, що завдання знецінюється пропорційно часу його перебування в системі, показник ефективності (2) — у припущенні, що при перевищенні деякого часу чекання завдання втрачає свою цінність відразу.

Більш загальним стосовно розглянутих показників є такий показник ефективності:

$$C_{T_{ВЗ}}^{(S)} = \sum_{i=1}^{n} \alpha_i' \lambda_i T_{ВЗi}^{(S)}, \tag{3}$$

де $T_{ВЗij} = \dfrac{\sum_i t_{ВЗij}}{N_i}$ — середня величина відносної затримки завдань *i*-го типу в системі;

$$t_{ВЗij} = \begin{cases} 0, & якщо \quad t_{ВИКij} \le T_{ДОПij} \\ \dfrac{t_{ВИКij} - T_{ДОПij}}{T_{ВТРij} - T_{ДОПi}} & якщо \quad T_{ДОПij} < t_{ВИКij} < T_{ДОПi} \\ 1, & якщо \quad t_{ВИКij} \ge T_{ДОПi}, \end{cases}$$

де $t_{ВЗij}$ — відносна затримка *j*-го завдання *i*-го типу в системі; $t_{ВИКij}$ — час перериванння *j*-го завдання *i*-го типу в системі від моменту надходження до моменту закінчення його обслуговування; $T_{ДОПij}$ — припустимий час переривання *j*-го завдання *i*-го типу в системі, при якій його цінність не знижується; $T_{ВТРij}$ — час, після закінчення якого *j*-го завдання *i*-го типу цілком знецінюється; N_i — кількість завдань *i*-го типу, що надійшли в систему за час *T*.

В основі показника ефективності (3) лежить припущення про те, що цінність завдання, яке надійшло до вузла, до деякого директивного часу $T_{ДОП}$ залишається постійною, а з моменту його перевищення з часом лінійно знижується до повного знецінювання в момент часу $T_{ВТР}$. Тому неважко показати, що при $T_{ДОПi} = T_{ВТРi} = D_i$ показник ефективності (3) приймає вигляд виразу (1), а при $T_{ДОПi} = 0$ і $T_{ВТРi} >> \upsilon_i$ — виразу (2). Використовуючи формули (1)–(3), можна оцінити ефективність *S*-го способу організації обчислювального процесу порівняно з *g*-м відношенням:

$$K^{(S,q)} = \frac{C^{(q)}}{C^{(S)}}. \qquad (4)$$

Недоліком показника ефективності (4) є труднощі у зіставленні виграшу, отриманого за допомогою *S*-го способу організації обчислювального процесу, з додатковими витратами обчислювальних ресурсів на реалізацію цього способу. Тому більш наочним і зручним показником для оцінки ефективності організації обчислювального процесу може служити виграш в еквівалентній продуктивності, сутність якого полягає в наступному. Нехай *S*-й спосіб організації обчислювального процесу зменшує величину сумарного штрафу порівняно з *q*-м способом у $K^{(S,q)}$ разів. Такий же результат можна одержати при *q*-му способі організації обчислювального процесу, збільшивши продуктивність вузла. Отже, показником ефективності організації обчислювального процесу може служити таке відносне збільшення продуктивності вузла та зменшення часу рішення всіх завдань $E^{(S,q)} = \dfrac{T_i^{(S)}}{T_i^{(q)}}$, при якій $C^{(q)}\{T_i^{(S)} / E^{(S,q)}\} = C^{(S)}\{T_i^{(S)}\}$, де T_i — час рішення *i*-го завдання у вузлі.

Найважливішим показником ефективності організації обчислювального процесу в системі є її фактична продуктивність:

$$\Psi = \sum_{i=1}^{N} u_i c_i, \tag{5}$$

де N — кількість процесорів у системі; c_i — швидкодія i-го процесора; u_i — коефіцієнт використання i-го процесора.

Для гомогенних середовищ (5) трансформується в сумарний коефіцієнт використання процесорів $G = \sum_{i=1}^{N} u_i$.

Однією з причин, що знижують продуктивність вузла, є виникнення конфліктних ситуацій, при яких два чи більше процесів одночасно вимагають один і той самий обчислювальний ресурс. Для оцінки впливу конфліктів на ефективність організації обчислювального процесу використовують наступні показники:

$$\delta = \left(1 - \frac{\psi}{\psi_0}\right) 100\ \%, \tag{6}$$

$$\delta' = \left(1 - \frac{G}{G_0}\right) 100\ \%, \tag{7}$$

де ψ_0, G_0 — відповідно фактична продуктивність вузла і сумарний коефіцієнт використання ресурсів без обліку конфліктних ситуацій.

Значення показника ефективності (6), (7) виражають втрату продуктивності (у відсотках) у вузлі, обумовлену виникненням конфліктних ситуацій.

Аналітичні моделі вузлів хмарних обчислень

Математичні моделі дозволяють дослідити різні режими організації обчислень у вузлах ХО. Для цього було виконано моделювання роботи вузлів з урахуванням впливу непродуктивних відволікань обчислювальних ресурсів. Кожен вузол моделюється системою масового обслуговування (СМО). На рис. 4 наведено структуру СМО, основними елементами якої є: вхідний потік заявок *A*; черга *Q*; дисципліна пріоритетного обслуговування *ДО*, що визначає порядок вибору заявок з черги; обслуговуючий пристрій *П*. До функцій системи входить: постановка заявок у чергу, вибір з черги заявки, що підлягає першочерговому обслуговуванню, і її обслуговування. На виході пристрою *П* утворюється вихідний потік *B*.

Для операційних вузлів характерні непродуктивні витрати обчислювальних ресурсів. Ці відволікання ресурсів у загальному випадку носять випадковий характер і в рамках теорії масового обслуговування, що вивчає системи і мережі масового обслуговування, їх можна інтерпретувати потоком відмов *ПВ* обслуговуючого пристрою, а їхню тривалість — часом його відновлення. Після відновлення обслуговуючого пристрою, що відмовив, обслуговування заявок починається відповідно до дисципліни відновлення *ДВ*. Відмови обслуговуючого пристрою викликають збільшення кількості необслугованих заявок, зростання черги заявок і додаткові затримки в їхньому обслуговуванні. У силу випадкового характеру процесу обслуговування заявок ці часові затримки для певних видів процесів можуть виявитися дуже значними, що матиме істотний вплив на

ефективність організації обчислювального процесу. Один із можливих способів адаптації системи до непродуктивних витрат обчислювальних ресурсів є впровадження відповідної дисципліни пріоритетного прийому заявок у чергу (*ДПВ*) від різних процесів на час відволікання ресурсів. Таке регулювання надходженням потоків може бути досягнуте за рахунок зворотного зв'язку ресурсу з джерелами заявок або шляхом закриття черги.

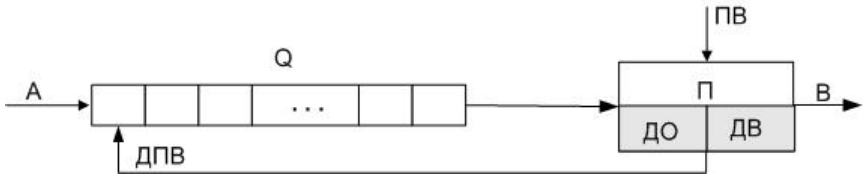

Рис. 4. Структура СМО, що моделює обчислювальний ресурс вузла

Задачу було сформульовано наступним чином. На вхід одноканальної СМО з очікуванням надходять N пуассонівських потоків різнотипних заявок з інтенсивністю λ_i, $i=\overline{1,N}$. Потоки перенумеровані в порядку зменшення важливості заявок, тобто заявки i-го потоку володіють i-м пріоритетом в обслуговуванні. Час обслуговування заявок є випадковою величиною з функцією розподілу $B_i(t)$ і двома скінченими моментами b_i і $b_i^{(2)}$, $i=\overline{1,N}$.

Обслуговуючий пристрій ненадійний і може виходити з ладу за законом Пуассона з параметром λ_0. Час відновлення пристрою — випадкова величина з функцією розподілу $B_0(t)$ і двома скінченими моментами b_0 і $b_0^{(2)}$. Пристрій може вийти з ладу як під час обслуговування заявок (при цьому можливі два випадки: заявки повертаються в чергу; заявки губляться), так і у вільному стані. У період відновлення обслуговуючого пристрою заявки одних потоків у чергу приймаються, а інші — не приймаються. Ця умова задається матрицею — рядком коефіцієнтів $n_i, i=\overline{1,N}$, де $n_i=1$ у тому випадку, якщо заявки i-го потоку в чергу приймаються, і $n_i=0$, якщо заявки одержують відмову в обслуговуванні.

Після відновлення обслуговуючого пристрою можливі дві дисципліни поновлення обслуговування: із заявок старшого пріоритету та із заявок, обслуговування яких було перервано відмовою пристрою (за умови, що вони не втрачаються протягом відмови). Потрібно визначити наступні характеристики обслуговування заявок: w_i — середній час чекання початку обслуговування заявок i-го потоку в i-й черзі; v_i — середній час перебування заявок i-го потоку в системі; q_i — середнє число заявок i-го потоку в i-й черзі; l_i — середнє число заявок i-го потоку в системі.

Сполучення дисципліни обслуговування з однією із дисциплін відновлення обслуговування після відновлення пристрою, що відмовив, і поведінкою заявки, обслуговування якої було перервано відмовою, визначило умови окремих задач для пріоритетних СМО. Розроблені аналітичні моделі для відносних, абсолютних, змішаних і комбінованих пріоритетних дисциплін дозволили отримати кінцеві вирази для шуканих характеристик обслуговування заявок.

Характеристики обслуговування заявок у сталому режимі зв'язані між собою формулами Літтла, що для систем із пріоритетним прийомом заявок до черги під час відно-

влення приладу, що відмовив, мають наступний вигляд:

$$l_i = \lambda_i^* \nu_i\,,\ q_i = \lambda_i^*\, w_i\,, \tag{8}$$

де $\lambda_i^* = K_r\lambda_i(1+n_i\rho_0)$ — інтенсивність надходження заявок *i*-го потоку в СМО з урахуванням дисципліни прийому до черги під час відновлення приладу; $K_r = \dfrac{1}{1+\rho_0}$ — імовірність того, що обслуговуючий прилад знаходиться в справному стані; $\rho_0 = \lambda_0 b_0$ — «завантаження» системи відмовленнями.

Умовою сталого режиму в системах даного класу без утрат є $\sum_{i=1}^{N} \rho_i^* < K_r$, де $\rho_i^* = \lambda_i^* b_i$ — імовірність зайнятості приладу обслуговуванням заявки *i*-го потоку.

Показники обслуговування вузлів хмарних обчислень

Наведемо отримані вирази для обчислення середнього часу очікування початку обслуговування для деякої заявки, що надходить у систему *j*-го потоку, $j = \overline{1, N}$, в явному вигляді для різних типів СМО, з яких неважко виводяться інші характеристики.

Система з відносними пріоритетами і поновленням обслуговування із заявок, обслуговування яких було перервано відмовленнями

$$\begin{aligned} w_j = \frac{1}{2(K_r - R_j)(K_r - R_{j-1})}\Big[& n_1 K_r^{\,2}\lambda_0 b_0^{(2)} + \sum_{i=1}^{N} \lambda_i b_i^{(2)}(1 + n_i\rho_0) + \\ & + K_r\lambda_0 b_0^{(2)} \sum_{i=2}^{j} (n_i - n_{i-1})(K_r - R_{i-1})\Big], \end{aligned} \tag{9}$$

де $R_j = \sum_{i=1}^{j} \rho_i^*$.

Система з відносними пріоритетами і поновленням обслуговування із заявок систем старшого пріоритету

$$\begin{aligned} w_j = \frac{K_r}{(K_r - R_j)(K_r - R_{j-1})}\Bigg\{ & n_1 K_r\sigma_0 + K_r\sum_{i=1}^{j}\sigma_i + K_r\sum_{i=j+1}^{N}\rho_i^*\Delta_i^* + \\ & + \sigma_0\sum_{i=2}^{j}(n_i - n_{i-1})(K_r - R_{i-1}) + \frac{1}{K_r}\sum_{i=2}^{j}[\rho_i^*\Delta_i^* - \sigma_i\overline{P_{ВІДМ}}(b_i)]R_{i-1}\Bigg\}, \end{aligned} \tag{10}$$

де $\sigma_0 = K_r\rho_0\Delta_0$; $\sigma_i = \rho_i(1 + n_i\rho_0)\Delta_i(1 + \rho_0)$; $\Delta_0 = \dfrac{b_0^{(2)}}{2b_0}$ — середній час до відновлення пристрою, що відмовив; $\Delta_i = \dfrac{b_i^{(2)}}{2b_i}$ — середній час дообслуговування заявки *i*-го потоку без обліку відмов обслуговуючого пристрою; Δ_i^* — середній час зайнятості

приладу дообслуговування заявки *i*-го потоку і відновленням після відмовлення, що наступило під час дообслуговування цієї заявки; $\overline{P_{ВІДМ}}(b_i) = 1 - \dfrac{P_{ВІДМ}(b_i)}{K_r}$.

Система з абсолютними пріоритетами:

$$w_j = \frac{K_r}{2(K_r - R_j)(K_r - R_{j-1})} \times \\ \times \left[n_1 K_r^2 \lambda_0 b_0^{(2)} + \sum_{i=1}^{j} \lambda_i b_i^{(2)} (1 + n_i \rho_0) + K_r \lambda_0 b_0^{(2)} \sum_{i=2}^{j} (n_i - n_{i-1})(K_r - R_{i-1}) \right]. \tag{11}$$

У системах зі змішаними пріоритетами сусідні потоки заявок поєднані у *M* груп, між якими діє абсолютний, а усередині кожної — відносний пріоритет в обслуговуванні. При цьому кожна *m*-та група потоків заявок містить у собі потоки з номерами від $(s_{m-1} + 1)$ до $s_m, m = \overline{1, M}$. На відміну від систем з абсолютними пріоритетами, заявка *j*-го потоку *m*-ї групи, що надходить на вхід системи, повинна очікувати в черзі дообслуговування заявок з потоків з номерами від 1 до s_m, і її обслуговування може бути перервано надходженням заявок старших пріоритетів у тому випадку, якщо $j > s_1$ (при цьому сумарний потік заявок, що переривають її обслуговування, включає потоки з номерами від 1 до s_{m-1}).

Тому для системи зі змішаними пріоритетами та поновленням обслуговування із заявок, обслуговування яких було перервано відмовленнями, рівняння для середнього часу чекання початку обслуговування набуде вигляду:

$$w_j^{(m)} = \frac{K_r}{2(K_r - R_j)(K_r - R_{j-1})} \left[n_1 K_r^2 \lambda_0 b_0^{(2)} + \sum_{i=1}^{s_m} \lambda_i b_i^{(2)} (1 + n_i \rho_0) + \right. \\ \left. + K_r \lambda_0 b_0^{(2)} \sum_{i=2}^{j} (n_i - n_{i-1})(K_r - R_{i-1}) \right]. \tag{12}$$

Так само зміниться вираз і для системи зі змішаними пріоритетами і поновленням обслуговування з заявок систем старшого пріоритету.

У системі з комбінованими пріоритетами час обслуговування всіх заявок, крім заявок першого потоку, розбито на два відрізки (етапи): на першому діє абсолютний пріоритет, на другому — відносний. Отже, тривалість першого етапу обслуговування заявки *k*-го потоку $z_{ik}; i = 1, k-1; k = \overline{2, N}$ — постійна величина, а на другому етапі тривалість обслуговування заявок залежить від дисципліни поновлення обслуговування. Так, для системи з комбінованими пріоритетами і поновленням обслуговування із заявок, обслуговування яких було перервано відмовленням, виразом для w_j в явному вигляді буде:

$$w_j = \frac{1}{2(K_r - R_j)(K_r - R_{j-1})} \sum_{i=1}^{j} [K_r^2 \lambda_0 b_0 (n_i - n_{i-1}) + \lambda_i (1 + n_i \rho_0) b_i^{(2)} + \sum_{k=i+1}^{N} \lambda_k (1 + n_k \rho_0) x_{ik}^{(2)} - \\ - \sum_{k=1}^{N} \lambda_k (1 + n_k \rho_0) x_{i-1,k}^{(2)} + \sum_{k=1}^{i-1} P_{ik} (2 b_k - z_{ik})](K_r - R_{i-1}). \tag{13}$$

За тим же принципом можна вивести рівняння для систем з комбінованими пріоритетами і поновленням обслуговування із заявок старшого пріоритету.

Відмінною рисою пріоритетних систем з утратами заявок є втрата заявок, обслуговування яких було перервано відмовою приладу. Тому ймовірність зайнятості приладу обслуговуванням заявки *i*-го потоку $\rho^*_{серi} = \lambda^*_i b_{серi}$, де $b_{серi}$ визначається за виразом: $b_{серj} = \int_0^\infty [1 - B_j(t)] e^{-\lambda_0 t} dt$, де $[1 - B_j(t)]$ — ймовірність того, що за час *t* заявка *j*-го потоку не буде обслугована пристроєм; а $e^{-\lambda_0 t}$ — імовірність того, що за час *t* не відбудеться жодної відмови. За аналогією із пріоритетними системами без утрат можна отримати в явному вигляді: вирази для w_j у системах з утратами заявок з відносними, абсолютними і змішаними пріоритетами в обслуговуванні.

На рис. 5 для двопріоритетної системи з відносними пріоритетами і режимом безперервного поповнення черги заявками вищого пріоритету протягом періоду відновлення пристрою проілюстровано залежність довжини черг заявок з нижчим (крива 1) та вищим (крива 2) пріоритетами від інтенсивності відмов обслуговуючого пристрою λ_0. Розрахунок проводився для таких значень параметрів:

$$\rho_1 = \rho_2 = 0{,}4;\ \mu_1 = \mu_2 = 4{,}44;\ \mu_0 = 0{,}1\ (a_1 = \mu_1 / \mu_0 = 44{,}4;\ a_0 = \mu_0 / \mu_2 = 0{,}0225).$$

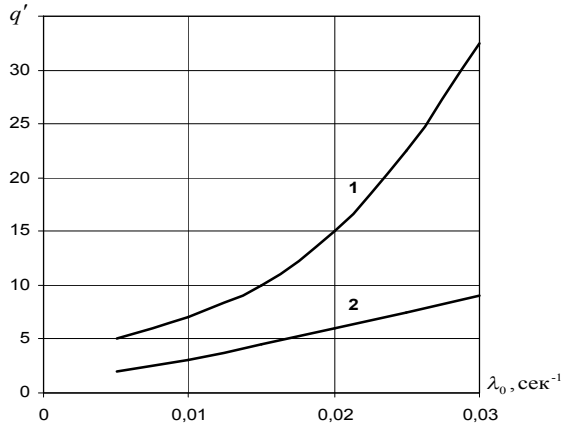

Рис. 5. Залежність довжини черг заявок (q') від інтенсивності відмов (λ_0)

Висновки

1. Технологія ХО надає якісно новий та економічний рівень інтеграції відомчих систем і їхніх ресурсів в інформаційно-аналітичну систему державного управління. Окрім цього, ця технологія поставила б на службу суспільству загальнодержавні та відомчі реєстри різного призначення (фізичних осіб, виборців і суб'єктів підприємницької діяльності, об'єктів нерухомості, власників паспортів, водійських прав тощо). А також вона надала б реальний інструментарій для створення єдиних інформаційних і довідкових служб і систем міст, регіонів, наприклад, з питань охорони здоров'я і екології, юридичних питань, внесення платежів, супроводу бюджету, казначейського обліку, загальнодержавного технагляду, митного контролю тощо.

2. У технологіях ХО можуть бути знайдені ефективні рішення і для завдань, що постали перед вітчизняним бізнесом: інтеграція мобільної телефонії з Інтернет, постачання телепрограм, відео, електронних підручників за замовленням, діловодства, бізнесу, торгівлі, логістики.

3. Розроблені математичні моделі дозволяють дослідити різні режими організації обчислень у вузлах ХО. Кожен вузол моделюється системою масового обслуговування з урахуванням впливу непродуктивних відволікань обчислювальних ресурсів. Наведені моделі для відносних, абсолютних, змішаних і комбінованих пріоритетних дисциплін обслуговування дозволили отримати кінцеві вирази для шуканих характеристик обслуговування. Сполучення дисципліни обслуговування з однією із дисциплін відновлення обслуговування визначило умови окремих задач пріоритетних систем. Кількісно спосіб організації обчислювального процесу може бути оцінений наведеними показниками ефективності через характеристики обслуговування.

1. *Carl Hewitt.* ORGs for Scalable, Robust, Privacy-Friendly Client Cloud Computing // IEEE Internet Computing. — 2008.— Vol. 12, N 5. — P. 96–99.

2. *Foster I.* The Anatomy of the Grid: Enabling Scalable Virtual Organizations [Електронний ресурс] / I. Foster, C. Kesselman and S. Tuecke // International Journal of High Performance Computing Applications. — 2001. — Vol. 15, N 3. — P. 200–222. — Режим доступу: www.globus.org/research/ papers/anatomy.pdf

3. *Матов О.Я.* Перспективні інформаційні технології та розвиток GRID-систем у високопродуктивних глобально-розподілених обчислювальних інфраструктурах корпоративної співпраці / О.Я. Матов, І.О. Храмова // Реєстрація, зберігання і оброб. даних. — 2004. — Т. 6, № 1. — С. 85–98.

4. *Amazon* Elastic Compute Cloud (Amazon EC2) [Електронний ресурс]. — Режим доступу: http://aws.amazon.com/ec2/

5. *Amazon* Simple Storage Service (Amazon S3) [Електронний ресурс]. — Режим доступу: http://aws.amazon.com/s3/

6. *Корпоративна* електронна пошта, документи та сайти Інтранету для уряду — Служби Google для уряду [Електронний ресурс]. — Режим доступу: http://www.google.com/apps/intl/uk/government/index.html

7. *Amazon* Virtual Private Cloud (Amazon VPC). — Режим доступу: http://aws.amazon.com/vpc/

8. *GFD*-I.080. Open Grid Services Architecture. Version 1.5. — 2006. — Режим доступу: http://www.ogf.org/documents/GFD.80.pdf

9. *Grid* — Інші національні Grid [Електронний ресурс] — Режим доступу: http://grid.kpi.ua/index.php?option=com_content&task=view&id=20&Itemid=49&lang=ua

10. *Graham S.* Building Web Services with Java: Making Sense of XML, SOAP, WSDL, and UDDI. — [2nd ed.] / Graham S., Davis D., Simeonov S. [et al.] // SAMS. — 2005. — 816 p. — ISBN-13: 978-0-672-32641-7.

11. *Матов А.Я.* Организация вычислительных процессов в АСУ / А.Я. Матов, В.Н. Шпилев, А.Д. Комов [и др.]; под ред. Матова А.Я. — Киев.— 1989. — 200 с.

12. *Липаев В.В.* Эффективность методов организации вычислительного процесса в АСУ / В.В. Липаев, С.Ф. Яшков. — М.: Статистика, 1975. — 255 с.

13. *Матов О.Я.* Сучасні технології інтеграції інформаційних ресурсів / О.Я. Матов, І.О. Храмова // Реєстрація, зберігання і оброб. даних. — 2009. — Т. 11, № 1. — С. 33–42.

14. *Храмова І.О.* Застосування сервісно-орієнтованих архітектур у процесах інтеграції інформаційних ресурсів / І.О. Храмова // Реєстрація, зберігання і оброб. даних. — 2009. — Т. 11, № 2. — С. 70–76.

15. *Матов О.Я.* Математичні моделі конфліктних утрат продуктивності системи посередників онтології загального використання в GRID-середовищі / О.Я. Матов // Реєстрація, зберігання і оброб. даних. — 2009. — Т. 11, № 3. — С. 18–25.

16. *Матов О.Я.* Проблеми горизонтальної інтеграції інформаційних ресурсів у багаторівневих організаційних структурах з динамічною конфігурацією / О.Я. Матов, І.О. Храмова // Реєстрація, зберігання і оброб. даних. — 2007. — Т. 9, № 3. — С. 88–97.

17. *Матов О.Я.* Динамічна інтеграція інформаційних ресурсів єдиної інформаційної інфраструктури ринку електроенергії / О.Я. Матов, І.О. Храмова // Функціонування та розвиток ринків електроенергії та газу: зб. наук. пр. Ін-ту проблем моделювання в енергетиці ім. Г.Є. Пухова. — 2006. — С. 93–98.

18. *Матов О.Я.* Моделі продуктивності операційних вузлів інформаційної інфраструктури корпоративних інформаційних систем в галузі електроенергетики / О.Я. Матов, І.О. Храмова // Інформаційні технології в енергетиці: зб. наук. пр. Ін-ту проблем моделювання в енергетиці ім. Г.Є. Пухова. — 2006. — С. 95–105.

19. *Матов О.Я.* Організація онтологій загального використання в інтегрованих інформаційних інфроструктурах підготовки даних для прийняття рішень / О.Я. Матов, І.О. Храмова // Функціонування та розвиток ринків електроенергії та газу: зб. наук. пр. Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова. — 2006. — С. 99–103.

20. *Матов О.Я.* Проблеми використання GRID-технології як базису інтеграції інформаційно-аналітичних ресурсів для підтримки процесів електронного урядування / О.Я. Матов, І.О. Храмова // Вісті Академії інженерних наук України. — 2005. — № 2 (25). — С. 82–89.

Надійшла до редакції 15.05.2010