

## РОЗРОБКА ФІЗИЧНИХ І ЛОГІЧНИХ МЕТРИК ДЛЯ ЗАДАЧІ БАГАТОКРИТЕРІАЛЬНОЇ ОПТИМІЗАЦІЇ ІНФОРМАЦІЙНОГО НАВАНТАЖЕННЯ ПРИ СТРУКТУРИЗАЦІЇ КОРПОРАТИВНОГО ЦЕНТРУ ДАНИХ

*Анотація:* В статті розглядаються питання дослідження та розробки фізичних і логічних метрик засобів обробки інформації в корпоративних прикладних мережах, що виникає при їх оцінюванні та виборі в задачі багатокритеріальної оптимізації інформаційного навантаження компонентів таких мереж. Такі метрики серверів БД та файлових серверів, систем документообігу та архівації даних, керованих мережевих пристроїв, що виконують функцію фільтрації на рівні додатків, тощо, враховують вимоги з точки зору безпеки та з урахуванням обмежень по управлінню при структуризації корпоративного датацентру.

*Ключові слова:* центр даних, багатобагатокритеріальна оптимізація.

**Предмет, об’єкт дослідження** – Проблема заключається в тому, що для опису характеристик програмно-апаратних компонентів, які використовуються в архітектурі прикладних мереж і в основному визначають вибір технології зберігання даних, критеріїв їх вибору застосовують різні набори даних – в цифровій формі, логічні дані типу “так/ні”, текстовий опис, які складають відповідні фізичні та логічні метрики цих компонентів для вирішення задачі багатокритеріальної оптимізації їх структури та параметрів. Провідні ІТ-компанії пропонують еталонні архітектури, які частково полегшують обґрунтований вибір архітектури корпоративної мережі та функціональності пристроїв системи.

На цей час авторам відомі такі метрики тільки для окремих компонентів на основі шаблонів, що пропонуються компанією Майкрософт [2] для параметричної ідентифікації цих компонентів (поза задачею багатокритеріальної оптимізації балансування інформаційного навантаження) для експертної оцінки їх вибору.

**Суть процесу дослідження** – виконано наукові дослідження по вибору фізичних і логічних метрик щодо основних параметрів задач балансування навантаження при заданих структурних обмеженнях, обумовлених відповідними вимогами до ризиків рівнів безпеки, яким відповідають зони і яруси в архітектурі прикладної мережі [2], з метою створення алгоритмів параметричної та структурної оптимізації системи БД та сховища електронних даних Датацентру університету.

**Основні науково-технічні результати роботи** – Проведено аналіз критеріїв, на основі яких визначаються базові характеристики основних програмно-апаратних компонентів Датацентру, їх функціональне наповнення в задачах обробки та зберігання даних, показана залежність отриманих оцінок від розташування по зонах і ярусах на основі обмежень

безпеки та управління. Визначені вектори фізичних і логічних метрик для систем зберігання даних в такій мережі. Розроблено відповідне програмне забезпечення для такої процедури оптимізації в режимі експертного діалогу.

Архітектура системи зберігання даних має відповідати обраній інформаційній структурі, щоб гарантувати, що організація прийняла правильне рішення. У цьому питанні йдеться про три основних типи архітектури системи зберігання даних [3]: розподільне зберігання даних, гібридне зберігання даних, централізоване зберігання даних.

Кожен із цих типів архітектури системи зберігання даних передбачає свій спосіб зберігання, який можна використати як відправну точку для надання рекомендацій щодо того, як інтегрувати систему зберігання даних із потребами організації.

Структурований процес проектування повного набору рішень, пов'язаних із системою зберігання даних підприємства, складається з таких етапів [2]:

- визначення вимог до системи зберігання даних;
- вибір технологій зберігання даних;
- визначення технологій забезпечення стійкості до перебоїв;
- визначення технологій резервного копіювання та відновлення.

Проектуючи архітектуру системи зберігання даних, дуже важливо знати різні вимоги до зберігання даних, наприклад, до їх доступності, ємності та продуктивності. Щоб зрозуміти ці вимоги, потрібно знати точні метрики та усвідомлювати важливість критеріїв збирання цих метрик. Оцінка – важливий аспект процесу проектування, а якість метрик безпосередньо позначається на здатності розробника зробити обґрунтовані оцінки [4].

Щоб визначити фізичні метрики, спочатку потрібно зрозуміти, яка ємність потрібна для операційної системи; таку інформацію можна отримати дуже просто. Важче оцінити вимоги до самої ємності різних прикладних програм. Якщо відомо, які прикладні програми застосовуватимуться у вашому середовищі, у вас буде базова інформація, але робити узагальнені припущення не завжди безпечно.

Як правило, обсяги даних у різних організаціях суттєво відрізняються. Наприклад, кількість даних у системах передачі повідомлень може змінюватися залежно від таких чинників, як політики щодо архівації електронних листів та використання вкладень до повідомлень, визначені в межах організацій, а також обсяги електронної пошти. Важливо також пам'ятати, що різні прикладні програми висувають різні вимоги до зберігання даних.

Фізичні метрики певною мірою є безпосередніми. Для їх отримання використовують засоби опитування серверів [2].

Логічні метрики стосуються вимог до ємності запам'ятовуваних пристроїв, тому деякі загальні критерії проектування можуть не підходити для них. Наприклад, безпека – це не функція ємності, хоча на неї

найбільше треба зважати під час вибору технологій зберігання даних. Основні логічні метрики стосуються доступності, масштабованості, керованості та продуктивності[3]. Кожна з них може безпосередньо вплинути на ємність пам'яті, а також на вимоги до пропускну здатності.

Наприклад, вимоги до доступності та прийнятого рівня пристроїв можуть означати, що ємність запам'ятовуючих пристроїв у інфраструктурі потрібно збільшити вдвічі у середовищах, де недоступність може коштувати мільйонів гривень на день, при проектуванні системи зберігання даних важливо передбачити ємності для зберігання копій наборів даних.

Масштабованість системи зберігання даних – це також важлива логічна метрика. Поточні вимоги до зберігання даних складаються на основі фізичних метрик, які можна обчислити, а майбутні вимоги завжди представлені оцінками. Здатність забезпечувати достатню для масштабування ємність записуючих пристроїв можлива за умови знання місця зберігання непостійних даних та вимог, які іноді висуває найнесприятливіший сценарій. (Проте треба пам'ятати, що проектування найнесприятливішого сценарію може дорого коштувати).

Керованість, як правило, менше впливає на вимоги до загальної ємності системи, проте це не стосується систем, які вимагають багатьох заходів моніторингу.

Продуктивність, по суті, частіше пов'язана з технологіями, ніж із проблемою зберігання даних. Суттєво впливають на продуктивність кількості дисків, а також їх конфігурація та взаємозв'язки між ними.

Вибір технологій зберігання даних зумовлюється кількома чинниками. Застосування логічних метрик у процесі вибору технології більше стосується саме технологій, а не конкретних виробів. Проте важливо забезпечити спільну роботу всіх технологій та переконатись, що операційна система підтримує конфігурації апаратного забезпечення. Хоча при цьому можна шукати обхідні шляхи та компромісні рішення, для створення доступного, захищеного, масштабованого та керованого середовища краще інтегрувати заземодоповнюючі технології.

В Табл. 1 та Табл. 2 наведено приклади розроблених фізичних та логічних метрик для системи зберігання і сховища даних прикладної мережі університету.

Табл. 1:

Фізичні метрики системи зберігання і сховища даних

Таблица 1

Загальні вимоги до файлів: фізичні метрики	Конкретні вимоги
<b>1 Архітектура системи</b>	
- кластер (локальний)	Підтримує/ні
- кластер (розподілений) багато вузловий	
а) однієї організації	Так/ні
б) кількох організацій	Перелік
в) кількох країн	Перелік

- суперкомп'ютерне сховище	Ємкість/технологія
- кількість процесорів (чи вузлів)	число
- програмно-апаратні засоби захисту	(брандмауер/прокси/маршрутизатори)
- технологія архівації	DAS/ NAS/ SAN
<b>2 Операційна система базова / допоміжна</b>	
- дисковий простір, який потрібен для роботи (Gb)	min / optim
- обсяг оперативної пам'яті	min / optim
- обсяг КЕШ-пам'яті	число
<b>3 Прикладна програма архівації</b>	
- розмір (в Мб)	число
- робочий простір дискової пам'яті	min / optim
- рівень розподіленості	
- потрібні БД	Перелік БД
- потрібні СУБД чи системи архівації	Перелік/марка
- бібліотеки і класифікатори для роботи програм	Перелік
<b>4 Дані</b>	
- метадані опису об'єктів архівації (в Мб)	число
- дані тимчасового зберігання (GB)	число
- дані довгострокового зберігання (GB)	число
- обсяг даних, які щоденно пересилаються по мережі, в тому числі, як реплікації / в інтерактивному режимі	Число/число (байти)
- дані, що доступні в локальному режимі обробки(GB)	число
- приріст даних за період (день/місяць/рік) (GB)	число
<b>5 Протоколи з'єднання</b>	
- ТСП/Р	Так
- НТТР(S)	Так/ні
- інші	Перелік

Табл. 2: Логічні метрики системи зберігання і сховища даних

Таблиця 2

<b>1 Доступність</b>	
- час робочого стану (%)	Число
- середній час обслуговування запиту до архіву	число
- кількість запитів, які не обслуговано	число
- кількість запитів, які передані для обробки на інші системи	
а) в середині організації	число
б) за межами організації	число
- кількість рівнів захисту	цифра

<b>Загальні вимоги до файлів: логічні метрики</b>	<b>Конкретні вимоги</b>
- кількість доступних БД та сховищ	число
- середня кількість завдань в черзі (за годину / за добу)	Число/ число
<b>2 Масштабування</b>	
- очікуваний приріст даних на наступний період довгострокового зберігання	
* 1-рік (Gb)	число
* 2-рік (Gb)	число
- очікуваний приріст даних, які щоденно пересилаються по мережі (Gb)	число
- очікуваний приріст запитів до системи архівації	число
- очікуваний приріст даних, які обробляються в локальному режимі обробки в розрізі:	
* структурних підрозділів (Gb)	число
* користувачів (Gb)	число
<b>3 Продуктивність</b>	
- чисельність активних користувачів	
* 1-й рік	число
* 2-й рік	число
- час роботи	
* в автономному режимі	%
* в режимі GRID /вузла	%
- кількість оброблених запитів за період	
* за годину	число
* за день	число
* за тиждень/місяць/квартал	Число/ число/ число
- кількість помилкових запитів	% від загального
- кількість підтримуємих БД / сховищ	Число/ число
- наявність аварійних планів (критичність)	
* при втраті даних	Так/ні
* при несанкціонованій зміні даних	Так/ні
* при крадіжці даних	Так/ні
<b>4 Супроводження</b>	
- очікувана довговічність системи (років)	число
- етапи модернізації (по підсистемах/ в часі)	перелік
- глибина модернізації (по рівнях)	Перелік рівнів
- зміна технології зберігання (умови)	Перелік умов
- зміна технології обробки (умови)	перелік умов
<b>5 Керованість</b>	
- через групові політики адміністратором безпеки	Так/ні
- локально адміністратором БД / сховища	Так/ні
<b>6 Безпека</b>	
- наявність Active Directory	Так/ні

<b>Загальні вимоги до файлів: логічні метрики</b>	<b>Конкретні вимоги</b>
- обмеження операцій над даними сховищ	Перелік по категоріям користувачів
- вимоги конфіденційності даних	Перелік категорій даних
- політики обміну даними (між вузлами / між системами)	Так/ні

Вимоги щодо доступності потрібно оцінювати окремо для кожної служби в процесі збирання вимог до зберігання даних. Отримана інформація підкаже, який тип технології відповідає вашим вимогам. Загалом для доступності велике значення мають такі чинники, як надлишковість, стійкість до перебоїв та реплікація даних. Доступність також впливає на конфігурації топології, особливо в складних SAN. Необхідність забезпечення високого рівня доступності означає, що складні конфігурації “ядро-межа” з використанням двох незалежних груп комутаторів (dual-fabric core-to-edge configurations), використовувати краще, ніж більш прості реалізації типу “точка-точка”. Конфігурації топології можуть також суттєво впливати на продуктивність.

Виходячи з наявності різних форматів опису окремих характеристик для фізичних і логічних метрик, не існує такої єдиної формалізації, що дозволяє обробити та узгодити таке параметричне різноманіття. Тому тільки експерт може орієнтуватися в такому середовищі та оцінювати означені характеристики і вести “правильний” діалог з персональним комп’ютером по обгрунтованому вибору таких необхідних параметрів.

Авторами розроблено двоетапну процедуру і створено відповідний програмний інструментарій, які дозволяють на базі розроблених метрик для основних серверів та інших апаратних засобів корпоративної мережі [5] оптимізувати архітектуру та обгрунтовано вибрати відповідні пристрої Датацентру, що найбільш функціонально оптимальні з точки визначених обмежень безпеки та управління, а також мінімізують фінансові витрати на їх реалізацію, що буде описано в наступному збірнику.

### **Література**

1. Алексеев В.А. Архитектура предприятия. – <http://www.INTUIT.ru>
2. Еталонні архітектури MSA.- К., Майкрософт Україна; К.: Видавнича група BHV, 2005.- 352с.
3. Еталонні служби MSA.- К., Майкрософт Україна; К.: Видавнича група BHV, 2005.- 912с.
4. Терехина А.Ю. Методы метрического шкалирования. – М.,Наука, 1984,-207с.
5. О.А. Стенін, Ю.А. Тимошин, Т.Г. Шемседінов. –К., Закл. Звіт про НДР держ. Реєстр. 0106U0002475, НТУУ “КПІ”,2008.-115с.

Отримано 07.12.2009 р.