

вихідної інформації / О.В. Лаврухін // Вісник Дніпропетровського національного університету залізничного транспорту імені академіка В.Лазаряна. – Дніпропетровськ, 2008. - Вип. 25. – С. 162 – 164.

2. Лаврухін О.В. Удосконалення технології розподілу вагонів на основі автоматизації процесів змінно-добового планування / О.В. Лаврухін // Вісник економіки транспорту і промисловості. – Харків, 2008. – Вип. 22. – С. 63-64.

3. Лаврухін О.В. Визначення цільової функції пріоритетного відправлення вантажних поїздів зі станції / О.В. Лаврухін // Восточно-Европейский журнал передовых технологий, 2011. - Вип. № 2/10 (50). – С. 20-25.

4. Лаврухін О.В. Визначення економічних параметрів цільової функції пріоритетного відправлення вантажних поїздів зі станції / О.В. Лаврухін // Восточно-Европейский журнал передовых технологий, 2011. - Вип. № 3/3. – С. 29-31.

Анотації:

Сформовано основні підходи щодо реалізації системи підтримки прийняття рішень оперативного управління з розподіленим штучним інтелектом

Ключові слова: оперативне планування, розподілений інтелект, інтелектуальні модулі, основні контури, дорожній контур.

Сформированы основные подходы по реализации системы поддержки принятия решений оперативного управления с распределенным искусственным интеллектом

Ключевые слова: оперативное планирование, распределенный интеллект, интеллектуальные модули, основные контуры, дорожный контур.

Formed the main approaches to implement decision support systems operational management of distributed artificial intelligence

Keywords: operational planning, distributed intelligence, intelligent modules, the basic contours, road circuit.

УДК 621.39

ЯРЕМКО І.М., старший викладач (ДонНТУ);
ДОЛГІХ І.П., старший викладач (ДонНТУ).

Оптимізація функціонування центру обробки даних

Постановка проблеми

Функціонування центру обробки даних (ЦОД) істотно залежить від складу завдань, вихідних даних і сигналів, які поступають на вхід ЦОД – робочого навантаження. Робоче навантаження залежить від призначення ЦОД і є різною для центрів, які оперують різними інформаційними потоками. Плануючи розміщення серверів в ЦОД, необхідно забезпечити найбільш по-

вну відповідність центра обробки до свого призначення. Сформувані набір характеристик таким чином, щоб вони давали найбільш повне представлення про ефективність функціонування ЦОД [1,2].

Постановка задачі

Функціонування центру обробки даних представляється у вигляді певного процесу. Процеси підрозділяються на

прикладні й системні. Прикладні реалізують основні функції, задані прикладними програмами або обробними програмами ОС, а ініціюються завданнями користувачів або сигналами, що надходять у систему зовні.

Процес P_i описується як $P_i = \{t_i, A_i, T_i\}$, де t_i – момент ініціювання процесу. A_i – атрибути процесу, що визначають обчислювальну складність, обсяг оброблюваних даних, вимоги до часу відгуку, частоту запитів і T_i – траса процесу.

Траса процесу – послідовність подій, пов'язаних зі зміною стану процесу. Траса процесу представляється у вигляді впорядкованої множини подій $T_i = \{S_1, S_2, \dots, S_M\}$, що мали місце в моменти часу t_1, \dots, t_M , причому

$t_1 \leq t_2 \leq \dots \leq t_M$. До подій відносяться моменти старту процесу, початку й завершення обробки окремих завдань, початку й закінчення виконання процесів у зовнішніх пристроях центру обробки даних, початку використання й звільнення розділів пам'яті, надаваних процесу в запам'ятовувальних пристроях, і ін.

Траса може бути представлена у вигляді часової діаграми, що зображує виконання програми процесором і пам'яттю (рисунок 1). Траса складається з періодів, коли процесор і пам'ять зайняті виконанням програми й інтервалів часу, протягом яких процес перебуває в стані очікування, тобто не обслуговується жодним пристроєм, очікуючи моменту його звільнення.

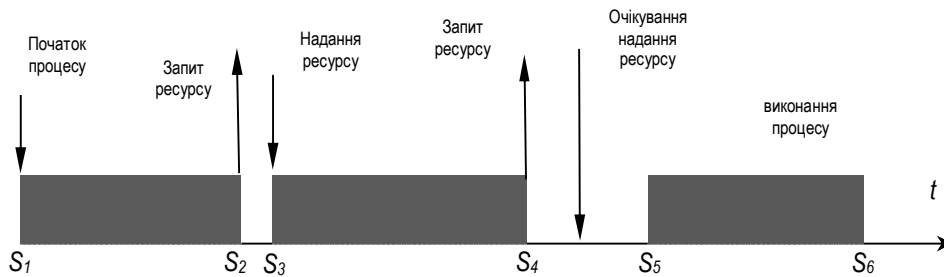


Рисунок 1. - Часова діаграма процесу в ЦОД

У розглянутій проблемній області найбільший інтерес представляє потреба завдань у ресурсах: оперативної й зовнішньої пам'яті, процесора часу. Тому робоче навантаження, яке відноситься до часу T , визначимо у вигляді множини характеристик завдань

$$\Lambda = \{\lambda_i\} = \{A_i, \Theta_{i1}, \dots, \Theta_{iN}\} \quad (1)$$

де λ_i , - опис і-го завдання, що встановлює його атрибути A_i і потреба завдання в ресурсах $\Theta_{i1}, \dots, \Theta_{iN}$.

Припустимо, що є потік вимог до ресурсу системи, усі вимоги поводяться статистично однаково, причому кожна з них вимагає деякого числа операцій від

ресурсу. Припустимо також, що є обчислювальна система (кластер серверів і систем зберігання даних) обслуговуюча M користувачів. Досліджуємо дану систему з погляду користувача, як показано на рисунку 2.

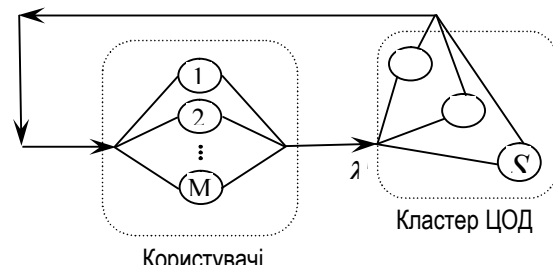


Рисунок 2. - Представлення ЦОД у вигляді мережі СМО

ЦОД як модель масового обслуговування

Скористаємося моделлю масового обслуговування замкненої мережі для системи колективного користування з багатьма ресурсами.

Нехай є замкнена мережа із загальним числом $K=M$ вимог, кожна з яких генерує вихідне від користувача вимогу з інтенсивністю λ вимог у секунду щораз, коли вимога перебуває в користувача, тобто всі вимоги характеризуються середнім часом міркування $1/\lambda$ секунд, яке виходить із довільного закону розподілу. Кожна, таким чином, сгенерована вимога надходить потім у ЦОД і переміщується від ресурсу до ресурсу відповідно до перехідних ймовірностей r_{ij} , зрештою вертаючись на термінал користувача. У цей момент розглянута вимога переходить до генерування нової вимоги.

Позначимо через T середній час проходження через ЦОД і через $1/\lambda$ - середній час перебування в користувача. Таким чином, середній час циклу дорівнює $T + 1/\lambda$, а середня пропускна здатність системи становить $\lambda' = M / (T + 1/\lambda)$ вимог у секунду. Аналогічно нехай N - середнє число вимог у ЦОД і M' - середнє число вимог у термінальному вузлі. На підставі формули Літтла маємо

$$T = \frac{N}{\lambda'} \quad (2)$$

Враховуючи, що $M = N + M'$, одержуємо

$$T = \frac{N}{\lambda'} = \frac{M}{\lambda'} - \frac{M'}{\lambda'} \quad (3)$$

Крім того, якщо застосувати формулу Літтла до термінального вузла, буде мати $1/\lambda = M'/\lambda'$, і, таким чином,

$$T = \frac{M}{\lambda'} - \frac{1}{\lambda} \quad (4)$$

Розглянемо насичений вузол, або вузьке місце в ЦОД, який на рисунку 2 позначений буквою S . Це такий вузол, у якому відношення x_s/m_s приймає найбільше значення з усіх x_i/m_i , у системі.

Щоб знайти x_i необхідно розв'язати систему рівнянь

$$\mu_i x_i = \sum_{j=1}^N \mu_j x_j r_{ji} \quad (5)$$

де $\mu_i x_i$ може бути інтерпретоване як відносне число відвідувань i -го вузла при циркуляції вимог у мережі. Для кожного розглянутого випадку $\mu_s x_s / \mu_N x_N$ - середнє число відвідувань вузького місця на кожне надходження вимог у кластер.

Щоб знайти число насичуючих вимог M^* , припустимо, що M^* повинне бути дорівнює максимальному числу ідеально спланованих завдань, що не дають ніяких взаємних перешкод. Для кожного з обслуговуваних m приладів у насиченому вузлі можна спланувати максимальне число завдань, яке дорівнює необхідному часу обслуговування завдання в одному циклі, поділеному на час обслуговування завдання в насиченому вузлі за цикл. Загальний час обслуговування за цикл дорівнює

$$\sum_{i=1}^N \frac{\mu_i x_i}{\mu_N x_N} \cdot \frac{1}{\mu_i}, \quad (6)$$

тобто сумі по всіх вузлах добутоків числа відвідувань вузла на цикл, на середній час обслуговування на відвідування.

Час перебування в насиченому вузлі дорівнює

$$\frac{\mu_s x_s}{\mu_N x_N} \cdot \frac{1}{\mu_s} \quad (7)$$

Тому що є m_s таких обслуговуючих приладів, відношення загального часу обслуговування за цикл вчасно перебування в насиченому вузлі за цикл може бути помножене на m_s що в підсумку дає

$$M^* = \frac{m_s}{x_s} \sum_{i=1}^N x_i \quad (8)$$

Заданою метою з'ясувати, як поліпшується характеристика системи в цілому, якщо усунути вузьке місце шляхом збільшення інтенсивності обслуговування в насиченому вузлі. Легко припустити, що в результаті цього який-небудь інший вузол, наприклад S' стане вузьким місцем системи.

Як показано в роботі [3], якщо продовжувати цю процедуру усунення вузьких місць, завжди можна чекати поява нового вузького місця.

Дана характеристика має важливе значення для визначення ефективності функціонування ЦОД, тому що дозволяє оцінити вигоду у поліпшенні характеристик ЦОД при використанні потужніших апаратних засобів і/або досконалих методів управління потоками інформації.

Висновки

Таким чином, завдання дослідження формулюється як завдання оптимізації функціонування замкненої системи масового обслуговування. Завдання оптимізації полягає в максимізації продуктивності ЦОД при вартості, що не перевершує задану

або мінімізації вартості ЦОД при продуктивності не нижче заданої.

Список літератури

1. В.Г. Олифер. Компьютерные сети: принципы, технологии, протоколы: Учебн. Для вузов 2-е изд. / В.Г. Олифер, Н.А. Олифер – СПб. Питер, 2005 – 864.
2. Безрук В.М. Методы многокритериальной оптимизации информационных систем/ В.М. Безрук // Радиоэлектроника и автоматика. – 1999. – Вып.2/07. с.63-68.
3. Л.Клейнрок. Вычислительные системы с очередями: Пер. с англ./ Л.Клейнрок, – М.: Мир, 1979.-600с.

Анотації

Ключові слова: центр обробки даних, оптимізація, система масового обслуговування.

Розглянуто постановку задачі оптимізації функціонування центру обробки даних. Функціонування центру обробки даних представляється у вигляді певного процесу, який описується кількома параметрами: моментом ініціювання, атрибутами і трасою. Розглянуто ЦОД як модель масового обслуговування замкненої мережі для системи колективного користування з багатьма ресурсами.

Рассмотрена постановка задачи оптимизации функционирования центра обработки данных. Функционирование ЦОД представляется в виде определенного процесса, который описывается несколькими параметрами: моментом инициации, атрибутами и трассой. Рассмотрены ЦОД как модель массового обслуживания замкнутой сети для системы коллективного пользования с многими ресурсами.

We consider the formulation of the problem of optimizing data center. Operation of the data center is in the form of a specific process, which is described by several parameters: the point of initiation and route attributes. Examined the data center as a model of a closed queuing network for shared-use system with many resources.