

# АНАЛІЗ ПОКАЗНИКІВ НАДІЙНОСТІ ТА ДОСТУПНОСТІ ВІДМОВОСТІЙКОГО КЛАСТЕРУ

Циліурік В.Є., Токар Л.О., Солоділов В.В., Муха Р.В.

Кафедра інфокомунікаційної інженерії ім. В.В. Поповського,  
Харківський національний університет радіоелектроніки,  
Україна.

E-mail: [vadym.tsyliuryk@nure.ua](mailto:vadym.tsyliuryk@nure.ua),  
[liubov.tokar@nure.ua](mailto:liubov.tokar@nure.ua),  
[rostyslav.mukha@nure.ua](mailto:rostyslav.mukha@nure.ua)

---

## Abstract

*The paper examines the advantages of high-performance HPC computing using a server cluster. It is shown that along with the advantages in the cluster environment, there are certain problems related to issues of system failures and achieving the maximum possible level of its availability. An overview of one of the basic elements for building cluster systems based on Asterisk was conducted. A solution based on a high-availability cluster using Asterisk with the possibility of providing high-reliability services is presented. Key reliability indicators, through which system availability and fault tolerance are evaluated, and SLA service level indicators used by providers to evaluate the guaranteed level of service availability are considered. The SLA indicators of well-known cloud providers, which are obliged to provide a certain percentage of availability time, were analyzed.*

---

Сучасний світ не можна представити себе без телекомунікацій - основної сфери, що забезпечує високу швидкість і ефективну передачу інформації. Інтернет та мобільні пристрої сприяли збільшенню потреб в комутаційних потужностях телекомунікаційних серверів. В останні роки спостерігається зростання інтересу до високопродуктивних обчислень High Performance Computing (HPC) з використанням кластера серверів. Завдяки цьому приватні особи отримали можливість ексклюзивного доступу до власних суперкомп'ютерів. Кластери на апаратному забезпеченні мають ряд переваг у порівнянні з традиційними суперкомп'ютерами.

По-перше, кластери з високою продуктивністю покликані стати дешевшою заміною більш складних та дорогих суперкомп'ютерів для виконання традиційних технічних додатків, таких як імітаційне моделювання, біотехнології, моделювання фінансових ринків, пошук даних та обробка потоків [1].

По-друге, кластерні обчислення можуть масштабуватися до великих систем. Сотні та навіть тисячі машин можуть бути об'єднані в мережу відповідно до потреб програми.

Третьою перевагою є доступність: заміна «несправного вузла» в кластері є тривіальною в порівнянні з усуненням несправності SMP-компоненти, що призводить до зниження середнього часу відновлення Mean Time To Repair (MTTR) при ретельно продуманій конфігурації кластера [2].

Тема роботи є актуальною, що обумовлено необхідністю дослідження параметрів роботи серверів у кластері для забезпечення потрібної відмовостійкості та мінімізації експлуатаційних витрат.

Однією з проблем у кластерному середовищі слід відзначити зведення до мінімуму відмов системи та досягнення максимально можливого рівня її доступності. Обчислення з високою доступністю High-Availability (HA) дозволяють уникнути проблем, пов'язаних з несподіваними відмовами, за рахунок активного резервування та запобіжних заходів.

Відмовостійкий кластер - група серверів, що гарантує мінімальний час простою віртуальних машин (ВМ). Відмовостійкі кластери використовують, наприклад, підтримки серверів баз даних, зберігання важливої інформації, роботи бізнес-додатків. Якщо один із серверів (вузлів) кластера втратив зв'язок з іншими вузлами або підключеним сховищем, VMmanager запустить процес аварійного відновлення - релокації ВМ:

- проведе перенесення VM з відмовленого вузла кластера на працюючий вузол;
- здійснить зупинку VM на відмовленому вузлі кластера;
- ізолює відмовлений вузол кластера.

Одним з базових елементів для побудови кластерів вважається Asterisk, який забезпечує безперервність бізнес-процесів і якість послуг на високому рівні. Asterisk є потужним і гнучким програмним забезпеченням з відкритим вихідним кодом, призначеним для створення телефонних станцій й систем обробки викликів.

Кластеризація Asterisk проводиться з метою підвищення рівня надійності вузлів телефонної мережі компанії, забезпечує підвищену стійкість до відмови телефонії, а також основних використовуваних сервісів. Рішення на базі кластера Asterisk та архітектура побудови таких систем дозволяє підвищити надійність відповідальних вузлів телефонії, що особливо стосується таких компаній, де важливим є кожен дзвінок клієнта, а також компаніям з великим потоком вхідних та вихідних дзвінків. Такий кластер вважається високодоступним.

Рішення на базі кластера високої доступності є телефонною платформою, побудованою на двох серверах, що працюють як єдина синхронізована система для надання високої надійності сервісів. Схема розподіленого незалежного кластера з використанням Asterisk показано на рис. 1.

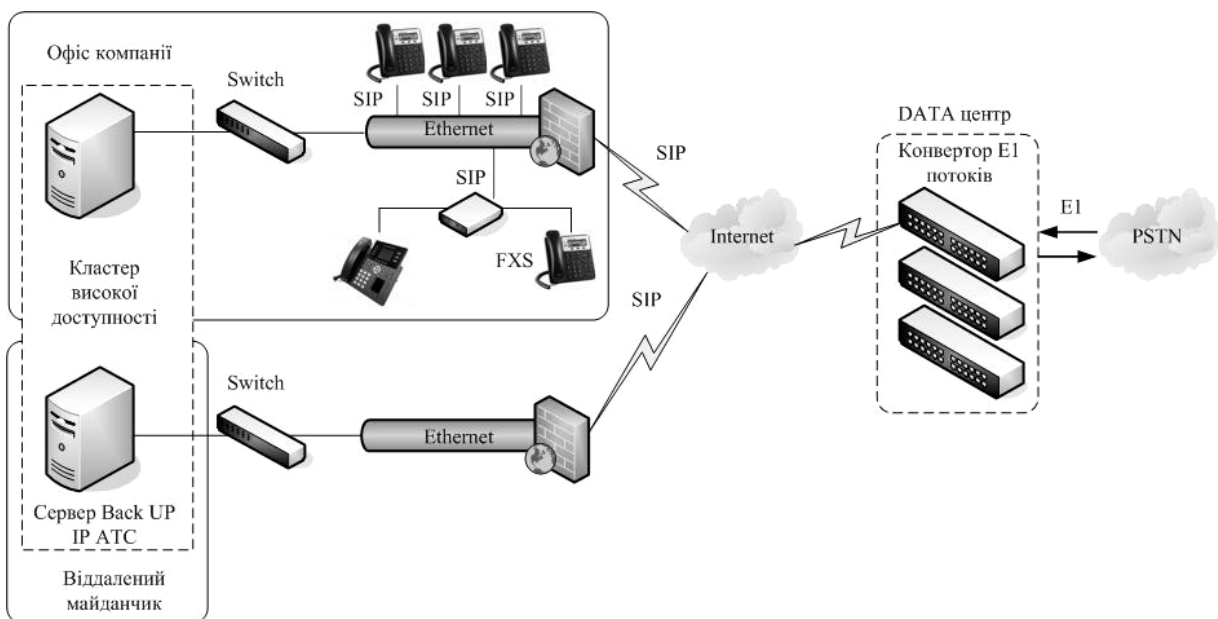


Рис. 1. Схема розподіленого незалежного кластера на базі Asterisk

Відмовостійкі рішення досягаються шляхом дублювання та синхронізації критично важливих для роботи телефонії сервісів та вузлів, а підвищена доступність ресурсів кластерної системи у разі збою одного із серверів забезпечується за рахунок переміщення служб або програм між вузлами в автоматичному режимі.

Однак, незважаючи на всі переваги, кластери з використанням Asterisk можуть стикнутися з проблемами доступності та відмовостійкості. Ці проблеми можуть бути пов'язані з обладнанням, програмним забезпеченням, мережними проблемами, людськими помилками та багатьма іншими факторами.

Доступність і відмовостійкість системи часто виражаються через ключові показники надійності, такі як середнім часом між відмовами Mean Time Between Failures (MTBF), середнім напрацюванням до першої відмови Mean Time To First Failure (MTTF) та середнім часом до ремонту Mean Time To Repair (MTTR) [3].

1. Середнім напрацюванням до першої відмови MTTF називається математичне очікування (момент першого порядку)  $M[t]$  часу роботи технічного засобу до відмови:

$$M[t] = T_{cp} = \int_{-\infty}^{+\infty} ta(t)dt \quad (1)$$

або

$$T_{cp} = \int_0^{+\infty} P(t) dt . \quad (2)$$

Фактично  $T_{cp}$  визначає інтервал часу, по закінченню якого ймовірність відмови складає:

$$T_{cp} = 1 - e^{-\lambda t} = 1 - e^{-1} = 0,63.$$

2. Середній час роботи між відмовами MTBF відновлюваного технічного засобу показує, яке напрацювання в середньому доводиться на одну відмову:

$$\overline{T_{cp}} = \sum_{i=1}^n \frac{t_i}{n} , \quad (3)$$

де  $t_i$  – час справної роботи технічного засобу між  $(i-1)$ -ою та  $i$ -ою відмовами;

$n$  – число відмов.

Для розрахунку MTBF використовується наступна залежність:

$$\overline{T_{cp}} = \frac{1}{\lambda} . \quad (4)$$

3. Характеристикою ремонтпридатності пристроїв є середній час їх відновлення MTTR. Між показниками надійності існують наступні співвідношення (рис. 2):

$$MTBF = MTTF + MTTR;$$

$$MTTF = MTBF - MTTR.$$

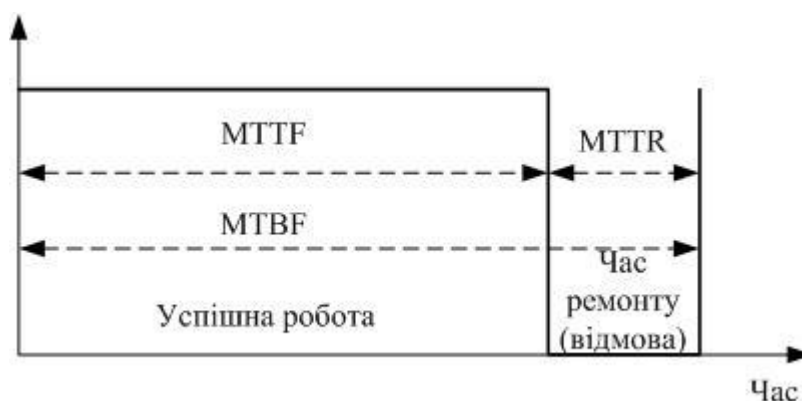


Рис. 2. Схематичне представлення MTTF, MTTR та MTBF

Доступність (A) системи можна розрахувати, використовуючи наступну формулу:

$$A = MTBF / (MTBF + MTTR)$$

Ця формула показує, яку долю часу система доступна для використання.

Відмовостійкість, з іншого боку, часто асоціюється з концепцією "п'яти дев'яток" (99,999% часу роботи), що часто використовується в індустрії телекомунікацій. Це означає, що система повинна бути доступна 99,999% часу, що відповідає приблизно п'яти хвилинам простою на рік.

Однак варто відзначити, що ці формули можуть бути узагальненими і не завжди точно відображають конкретні умови або конфігурацію спроектованої системи Asterisk. Ефективність і надійність системи також будуть залежати від багатьох факторів, включаючи архітектуру системи, якість обладнання та програмне забезпечення, а також процедури обслуговування та підтримки.

Показники рівня обслуговування Service Level Agreement (SLA) використовуються постачальниками публічних послуг для вказівки гарантованого рівня доступності їх послуг. Зазвичай вони забезпечують заздалегідь визначений рівень доступності у відсотках.

Ось як SLA може бути включено у вищевказану формулу доступності. Якщо припустити, що SLA - це рівень доступності, гарантований хмарним провайдером, у процентах (наприклад, 99,95%), то при перетворенні SLA в десятичну форму маємо:

$$SLA\_decimal = \frac{SLA}{100}, \text{ (наприклад, } 99,95\% \text{ буде } 0,9995).$$

Показник  $SLA\_decimal$  використовується для підрахування максимально допустимого часу простою в році. Рік зазвичай вважається рівним 525600 хвилинам (365 днів \* 24 години \* 60 хвилин). Максимальний час простою можна вичислити таким чином:

$$\text{Час простою} = (1 - SLA\_decimal) \cdot 525600.$$

Таким чином, використовуючи SLA постачальника зовнішніх послуг, можна підрахувати максимально допустимий час простою в хвилинах за рік.

Однак це дуже загальний підхід, і конкретні умови SLA можуть змінюватися в залежності від постачальника зовнішніх послуг. Деякі зовнішні провайдери можуть також надавати компенсацію за невиконання SLA, і ці умови також слід враховувати при оцінці відмовостійкості та доступності системи.

Угоди про рівень обслуговування SLA зовнішніх провайдерів відрізняються, але зазвичай вони зобов'язані забезпечувати певний відсоток часу доступності. В роботі проведено підсумки за місяць показників SLA відомих хмарних провайдерів.

1. Amazon Web Services (AWS): Amazon EC2 і Amazon EBS пропонують доступність SLA 99,99% для AWS кожного регіону. Якщо доступність падає нижче цього рівня, клієнти можуть отримати право на кредити у вигляді сервісних послуг.

2. Google Cloud Platform (GCP): Google Compute Engine забезпечує доступність SLA на 99,99% для всіх типів віртуальних машин, які запущені в двох або більше зонах. У разі недотримання цього зобов'язання Google може запропонувати кредити для майбутніх послуг.

3. Microsoft Azure: для більшості своїх сервісів, включаючи віртуальні машини та базу даних SQL, Microsoft забезпечує доступність SLA на 99,99%. У разі недотримання цих зобов'язань компанія може запропонувати кредити.

4. IBM Cloud: IBM обіцяє доступність за угодою SLA 99,99% для своїх сервісів, включаючи вичерпні сервіси та хмарне сховище.

Таким чином, рівні SLA та їх додаткові умови можуть відрізнятися залежно від конкретного сервісу та регіону, і вони можуть бути оновленими постачальниками зовнішніх послуг. Тому завжди рекомендується перевіряти актуальні дані на сайтах постачальників публічних послуг.

В роботі розглянуто переваги високопродуктивних обчислень НРС з використанням кластера серверів. Показано, що разом з перевагами у кластерному середовищі існують певні проблеми, пов'язані з питаннями відмов системи та досягнення максимально можливого рівня її доступності. Проведено огляд одного з базових елементів для побудови кластерних систем на основі Asterisk. Наведено рішення на базі кластера високої доступності з використанням Asterisk з можливістю надання високої надійності сервісів. Розглянуто ключові показники надійності, через які оцінюються доступність і відмовостійкість системи, та показники рівня обслуговування SLA, що використовуються постачальниками для оцінки гарантованого рівня доступності послуг. Проаналізовано показники SLA відомих хмарних провайдерів, які зобов'язані забезпечувати певний відсоток часу доступності.

## Література

1. Boukerche A., Al-Shaikh R. A. & Notare M. S. M. A. Towards highly available and scalable high performance clusters. *Journal of Computer and System Sciences*. 2007. Vol. 73(8). P. 1240–1251.
2. Mugler J., Naughton T., Scott S., Leangsuksun C. OSCAR clusters. *Proceedings of the Linux Symposium*, Ottawa, Canada, 2003. P. 387 – 397.
3. Манко Г.І., Тараненко Ю.К., Тітова О.В., Трішкін В.Я., Швачка О.І., Чумаков Л.Д. Розрахунки систем контролю та керування: навчальний посібник / за ред. Манко Г.І. Дніпро: УДХТУ, 2018. 191 с.