Високонадійна програмно-апаратна платформа для грід-сервісів

Борецький О.Ф., Савченко А.І., Сальніков А.О., Слюсар Є.А., Судаков О.О., Чередарчук А.І., Бойко Ю.В.

Київський національний університет імені Тараса Шевченка, Інформаційно-Обчислювальний центр, Київ, Україна

grid@grid.org.ua

Анотація.

Ефективне використання ресурсів грід-мережі потребує забезпечення стабільної та безперебійної роботи базових координуючих сервісів грід-інфраструктури. Інтеграція до таких європейський проектів як WLCG та EGI накладає додаткові вимоги до показників доступності та надійності провайдерів ресурсів грід-інфраструктури. У даній роботі запропоновані методики підвищення відмовостійкості роботи грід-сервісів шляхом використання апаратно-програмної системи віртуалізації з автоматичною міграцією та системи моніторингу працездатності грід-сервісів.

Ключові слова

Грід, віртуалізація, висока доступність, oVirt, Monit

1 Вступ

Одним з ефективних інструментаріїв вирішення ресурсоємних обчислювальних задач є грід-обчислення, що на сьогодні використовуються науковцями для досліджень в різних галузях науки і техніки.

Основними елементами грід-інфраструктури є обчислювальні елементи, елементів зберігання даних та грід-сервіси рівня кооперації, що координують роботу інфраструктури в цілому та є єдиною точкою входу для кінцевого користувача [1].

В Україні активно впроваджують і розвивають грід для вирішення освітніх та наукових задач з використаннях ресурсів провідних географічно-розподілених бюджетних установ. Так в 2006 році було створено Український академічний грід який, завдяки успішному та ефективному використанню, еволюціонував в Український національний грід (УНГ) [2].

На разі до складу УНГ входить понад 30 кластерів провідних науково-дослідних інститутів академії наук та університетів України[3]. Прозорий доступ до ресурсів кластерів грід-мережі залежить від коректної та стабільної роботи більшості елементів грід-інфраструктури.

Інтеграція грід-інфраструктури УНГ до таких європейських проектів як EGI та WLCG накладає ще більше вимог до забезпечення надійності та стабільності роботи координуючих грід-сервісів. Проблема розробки ефективних методик забезпечення високих показників доступності та надійності грід-сервісів є актуальною для функціонування як грід-інфрастурктури в цілому, так і її окремих складових.

В даній роботі запропоновані методики, що ґрунтуються на використанні програмно-апаратних засобів віртуалізації та моніторингу роботи сервісів, які забезпечують безперебійну роботу провайдерів ресурсів грід.

2 Використання віртуалізації для розгортання грід-сервісів

Насьогодні для розгортання грід-сервісів, зокрема в УНГ, використовуються сервери під керуванням операційної системи сімейства GNU/Linux. При розгортанні грід-сервісів користуються принципом “один сервіс – один сервер”, згідно якого на одному фізичному сервері працює тільки одна грід-служба. Такий підхід дозволяє забезпечити стабільну роботу, зменшити вплив сторонніх програмних компонентів та значно підвищити загальну безпеку всієї системи.

Однак розгортання грід-сервісів з використанням принципу “один сервіс – один сервер” потребує великої кількості апаратних ресурсів, що в умовах обмеженого фінансування грід-проектів УНГ не завжди є можливим. Використання віртуальних машин замість фізичних дозволяє більш ефективно використати наявні ресурси. Один вузол кластера середньої потужності дозволяє задовольнити потреби грід-сайту в віртуальних машинах для розгортання грід-сервісів.

Також використання віртуалізації дозволяє отримати ряд переваг пов’язаних з адмініструванням грід-сервісів:

- керованість;
- масштабованість;
- гнучкість.

Проте використання віртуалізації на одному апаратному вузлі створює єдину точку збою – апаратний збій сервера впливає на доступність одразу всіх грід-сервісів.

Поєднання переваг відокремлення апаратних платформ в ідеології “один сервіс – один сервер” та керованості і ефективності віртуальних машин забезпечується програмно-апаратними засобами віртуалізації, що представляють собою окрему цілісну інфраструктуру.

Програмно-апаратна платформа віртуалізації, що забезпечує надійну роботу сервісів поєднує в собі:

- множину серверів з апаратною віртуалізацією під керуванням гіпервізора;
- мережу зберігання даних віртуальних машин – Storage Area Network (SAN);
- систему керування платформою віртуалізації, що здійснює автоматичну міграцію сервісів в залежності від стану апаратних компонентів.

3 Система керування програмно-апаратною платформою віртуалізацією oVirt

Ключову роль у реалізації високонадійної інфраструктури відіграє система управління програмно-апаратною платформою віртуалізації. Одним з найбільш функціональних рішень, що включає автоматичну систему міграції віртуальних машин і розповсюджується з відкритим вихідним кодом є система oVirt. oVirt це безкоштовна платформа управління віртуалізації через web-інтерфейс, що розробляється за підтримки Red Hat і є базовою системою віртуалізації Red Hat Enterprise Virtualization [4]. Робота oVirt спирається на використання бібліотеки libvirt, що дозволяє керувати віртуальними машинами на будь-якому гіпервізорі, що підтримується, зокрема, KVM, Xen та VirtualBox.

3.1 Архітектура oVirt

Система керування oVirt складається з наступних основоположних компонент:

1. oVirt-engine – управляючий компонент системи. Дозволяє, за допомогою веб-інтерфейсу, керувати вузлами віртуалізації сховищами даних та віртуальними машинами. Здійснює моніторинг стану всієї системи та, при необхідності, переносить віртуальну машину з одного апаратного сервера віртуалізації на інший. Автоматично здійснює балансування навантаження між вузлами віртуалізації.
2. Вузол віртуалізації – вузол на якому запущені сервіси Virtual Desktop and Server Management (VDSM). VDSM виступає інтерфейсом керуванню віртуальними машин за допомогою бібліотек libvirt для oVirt-engine. Керування віртуальними машинами здійснюється за допомогою XML-інтерфейсу виклику віддалених процедур (Remote Procedure Call). Надає доступ користувачам до віртуальних машин по протоколам SPICE та VNC.

3. Спільне сховище образів віртуальних машин – забезпечує запуск будь-якої машини віртуалізації на будь-якому вузлі віртуалізації.

Схематично взаємодія компонентів зображена на рисунку 1 [5].

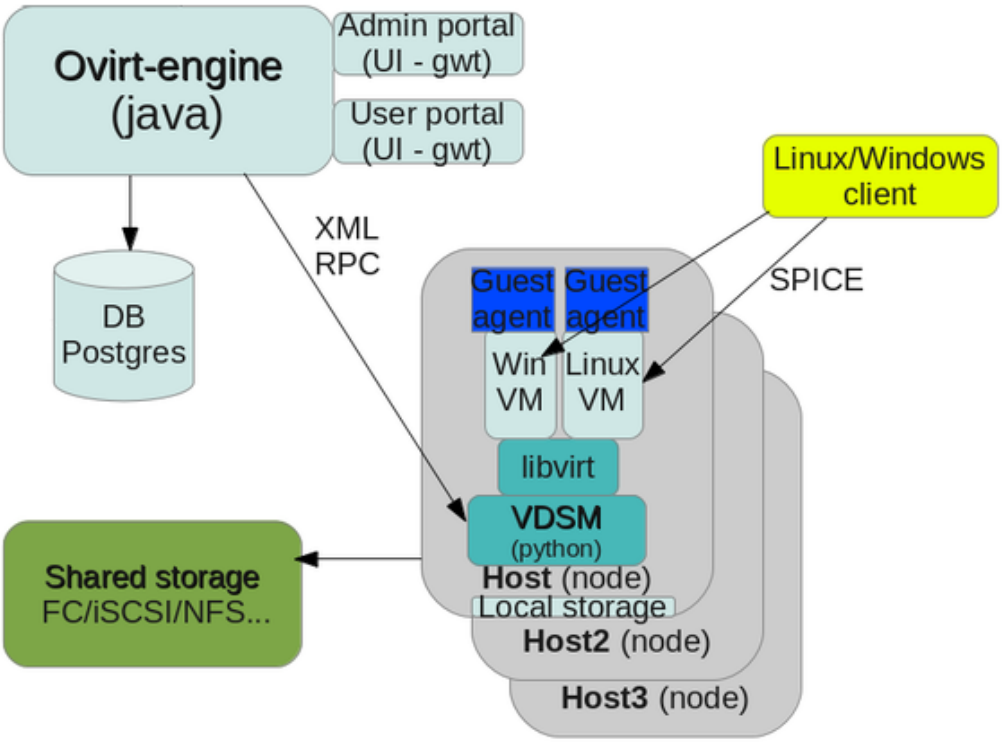


Рис. 1. Взаємодія компонентів системи oVirt

Систему керування апаратно-програмною платформою віртуалізації OVirt було впроваджено для розгортання координуючих грід сервісів та сервісів рівня ресурсів обчислювального кластеру [6] Київського національного університету імені Тараса Шевченка (КНУ).

3.2 Спільне сховище образів віртуальних машин

Одним із факторів завдяки якому досягається висока доступність віртуальних машин є їхня міграція між серверами віртуалізації. Однак для підтримки міграції необхідно забезпечити доступ до образів віртуальних машин з кожного сервера віртуалізації. Використання спільної мережі зберігання даних, доступ до якої з кожного сервера забезпечено на апаратному рівні дозволяє збільшити показники доступності.

Для побудови апаратного SAN на ресурсах обчислювального кластеру КНУ використано апаратну платформу зберігання даних HP EVA 3000, що надає доступ до образів віртуальних машин за допомогою технології Fiber Channel (FC). Система SAN складається з двох контролерів HSV100, трьох дискових масивів сумарним об'ємом 6 ТБ та двох комутаторів FC. Використання двох контролерів та комутаторів FC дозволяє гарантувати відмовостійкість сховища даних у системі віртуалізації.

На випадок перебоїв живлення SAN обладнано окремим блоком безперебійного живлення (UPS). Додатково кожен контролер HSV100 має 2 блоки свинцево-кислотних батарей для енергозалежної кеш-пам'яті, що зменшує імовірність втрати даних. Повністю заряджені батареї забезпечують живлення енергозалежної кеш-пам'яті на 96 годин [7].

3.3 Конфігурація мережі

В системі використовується дві окремі мережі:

- мережа для доступу до сховища образів віртуальних машин (SAN);

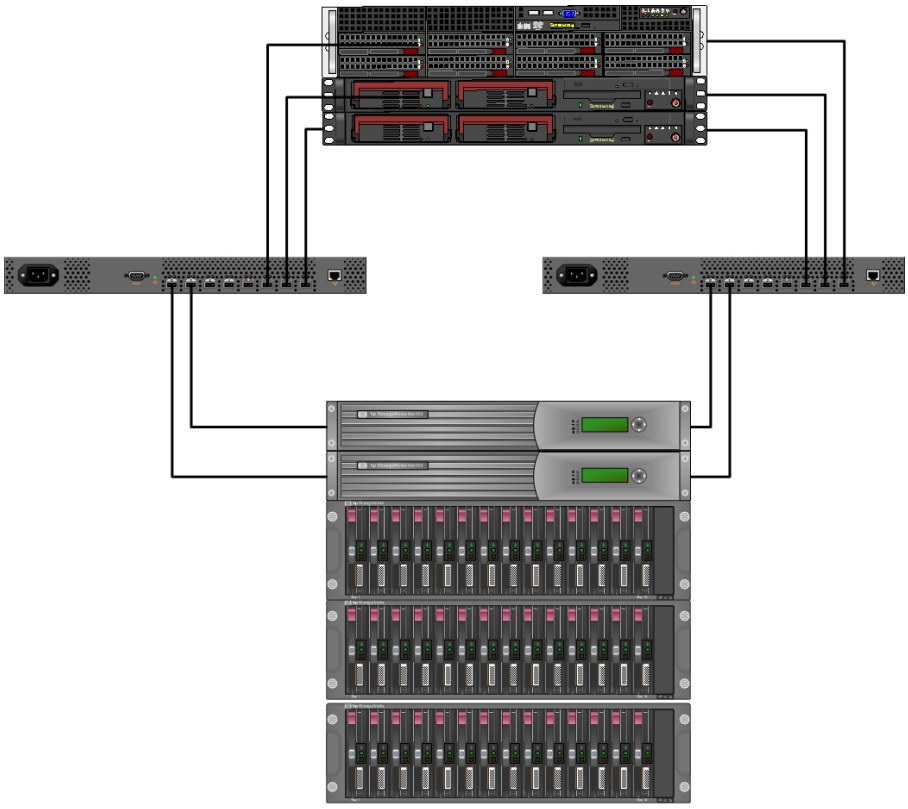


Рис. 2. Схематичне зображення мережі доступу до образів віртуальних машин

- мережа для підключення віртуальних машин до локальної мережі ресурсного центру і доступу до мережі інтернет.

Використання окремих мереж дозволяє більш ефективно розподілити навантаження в мережевій інфраструктурі, підвищити безпеку, зменшити вплив одна на одну.

Мережа доступу до спільного сховища образів побудована на основі технології Fibre Chanel. При побудові мережі було проведено дублювання каналів передачі даних. Кожний вузол віртуалізації обладнано FC-адаптером з двома портами та під'єднано до різних комутаторів. У такій конфігурації при виході з ладу контролера сховища даних, свіча або будь-якого каналу передачі даних мережа залишиться працездатною. Схематично топологія зображена на 2.

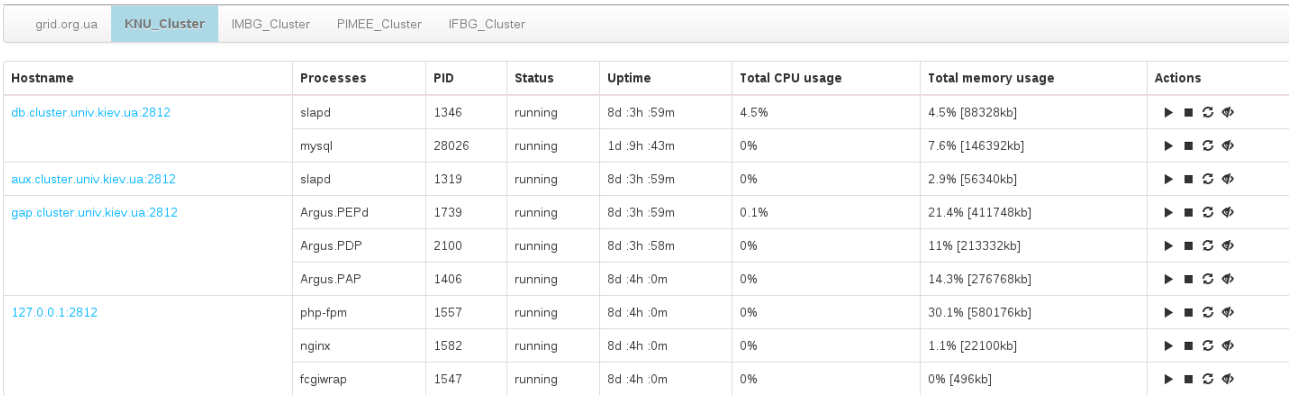
Віртуальні машини підключені до локальної мережі кластеру та мережі інтернет з використанням технології Gigabit Ethernet.На кожному сервері віртуалізації мережеві інтерфейси об'єднані за допомогою програмного комутатора. Завдяки такій конфігурації можна підключати велику кількість віртуальних машин через один фізичний інтерфейс.

Для розділення однієї фізичної на декілька логічних використано технологію VLAN. Зокрема створено окрему VLAN мережу для:

- системи віртуалізації oVirt;
- доступу до локальної мережі кластера;
- доступу до мережі інтернет.

4 Система моніторингу доступності грід-сервісів

Система керування програмно-апаратною платформою віртуалізації дозволяє забезпечити високу надійність та доступність елементів грід-інфраструктури на рівні віртуальних машин на яких працюють грід-сервіси. Однак повноцінна високонадійна програмно-апаратна платформа потребує інструменту для моніторингу та керування,



Hostname	Processes	PID	Status	Uptime	Total CPU usage	Total memory usage	Actions
db.cluster.univ.kiev.ua:2812	slapd	1346	running	8d :3h :59m	4.5%	4.5% [88328kb]	
	mysql	28026	running	1d :9h :43m	0%	7.6% [146392kb]	
aux.cluster.univ.kiev.ua:2812	slapd	1319	running	8d :3h :59m	0%	2.9% [56340kb]	
gap.cluster.univ.kiev.ua:2812	Argus.PEPd	1739	running	8d :3h :59m	0.1%	21.4% [411748kb]	
	Argus.PDP	2100	running	8d :3h :58m	0%	11% [213332kb]	
	Argus.PAP	1406	running	8d :4h :0m	0%	14.3% [276768kb]	
127.0.0.1:2812	php-fpm	1557	running	8d :4h :0m	0%	30.1% [580176kb]	
	nginx	1582	running	8d :4h :0m	0%	1.1% [22100kb]	
	fcgiwrap	1547	running	8d :4h :0m	0%	0% [496kb]	

Рис. 3. Централізована система моніторингу грід-сервісів

безпосередньо, елементами грід-сервісів, якими виступають Linux-демони. У роботі грід-сервісів можуть виникати внутрішні збої які пов'язанні з проблемами програмної реалізації чи недостатньої кількістю апаратних ресурсів, що не була виявлена завчасно.Такі збої можуть призводити до некоректної роботи процесу (наприклад, припинення обробки запитів) або до завершення його роботи. Враховуючи це система моніторингу доступності грід-сервісів має включати в себе сервіс по відновленню роботи демонів.

4.1 Monit

У якості базового елемента системи моніторингу грід-сервісів, що задовольняє необхідний функціонал, обрано Monit [8]. Monit – це легковажна система моніторингу Linux-демонів, яка має ряд переваг у порівнянні з аналогічними рішеннями:

- моніторинг стану серверів (доступність, використання ресурсів);
- моніторинг демонів (стан, використання ресурсів, кількість дочірніх процесів);
- моніторинг мережевих сервісів (наявність підключення і коректність відповіді);
- виконання вбудованих або власних (за допомогою сценаріїв) дій при виникненні визначених подій;
- надсилання сповіщень на електронну пошту.

Однак Monit здійснює моніторинг сервісів розташованих на одному сервері, а в умовах розподіленої грід-інфраструктури необхідно проводити моніторинг десятків грід-сервісів на різних серверах. Для вирішення такої задачі було створено власну реалізацію централізованого інтерфейсу для системи моніторингу грід-сервісів яка дозволяє, використовуючи централізований веб-інтерфейс керування проводити моніторинг та керувати всіма екземплярами Monit на різних віртуальних машинах разом зі сконфігурованими грід-сервісами. Характерною особливістю централізованого інтерфейсу є можливість групувати грід-сервіси в залежності від провайдера ресурсів (Рис. 3).

Систему моніторингу грід-сервісів було впроваджено в роботу програмно-апаратної платформи віртуалізації грід-сервісів Київськго національного університету імені Тараса Шевченка та для керування сервісами інших грід-кластерів УНГ.

5 Висновки

Проведено аналіз особливостей розгортання грід-сервісів рівня ресурсів та кооперації в умовах Української національної грід-інфраструктури. Запропоновано підходи до використання віртуалізації у ролі платформи для розгортання грід-сервісів з забезпеченням відмовостійкості до збоїв апаратних ресурсів.

Сформульовані вимоги до інфраструктури віртуалізації що дозволяють підвищити показники надійності та доступності віртуальних машин. Запропоновано методику побудови високонадійної інфраструктури віртуалізації для грід-сервісів на базі технлогій Fibre Channel та апаратної віртуалізації з використанням бібліотек Libvirt.

Для моніторингу та керування грід-сервісами в географічно розподіленій інфраструктурі було розроблено підходи до забезпечення доступності служб, які були втілені в програмному рішенні.

Методики та програмні засоби були успішно впроваджені в роботу обчислювального кластеру Київського національного університету імені Тараса Шевченка, що наразі є одним з основних провайдерів сервісів рівня кооперації, які забезпечують роботу УНГ вцілому.

Розроблені методики можуть бути застосовані для побудови інфраструктури віртуалізації та моніторингу грід-сервісів на ресурсах інших учасників УНГ.

Література

[1] Foster, Ian. The Anatomy of the Grid – Enabling Scalable Virtual Organizations / Ian Foster, Carl Kesselman, Steven Tuecke *International Journal of Supercomputer Applications*. — 2001. — Vol. 15. — P. 2001

[2] 2. Бойко Ю.В. Український академічний Грід: досвід створення й перші результати експлуатації / Ю.В.Бойко, М.Г.Зинов'єв, О.О.Судаков, С.Я.Свістунов *Математичні машині і системи*. — 2008. — Vol. 1. — Pp. 67–84.

[3] Український національний. – Online. URL: грід http://ung.in.ua/ua/. – 2013

[4] oVirt virtualization management platform. – Online. URL: http://www.ovirt.org – 2013

[5] oVirt Architecture. – Online. URL: http://www.ovirt.org/Architecture – 2013.

[6] Кластер Київського національного університету імені Тараса Шевченка. – Online. URL: http://cluster.univ.kiev.ua/ukr/ – 2013

[7] HP StorageWorks EVA 3000 — HSV100 Controller. — Online. http://h20000.www2.hp.com/bizsupport — 2013

[8] Process supervision tool Monit. – Online. URL: http://mmonit.com/monit – 2013