



Nghiên cứu xây dựng hệ thống tính toán hiệu năng cao phục vụ triển khai các bài toán trong Đại học thông minh

TRƯƠNG VIỆT PHƯƠNG^{a,*}, NGUYỄN QUỐC HÙNG^a, LÊ NGỌC THANH^a, VÕ HÀ QUANG ĐỊNH^a

^a Trường Đại học Kinh tế TP. Hồ Chí Minh

THÔNG TIN	TÓM TẮT
Ngày nhận: 30/01/2021 Ngày nhận lại: 29/05/2021 Duyệt đăng: 03/06/2021 Mã phân loại JEL: C61; C63; C67 Từ khóa: Tính toán hiệu năng cao; Xử lý cụm máy tính; Công nghệ Docker; Đại học thông minh; vHPC.	Chuyển đổi số trong giáo dục là một chủ đề đang thu hút được nhiều sự quan tâm của các nhà khoa học, nhà nghiên cứu trong thời gian gần đây. Công việc này giúp cho việc tin học hóa toàn bộ quá trình quản lý giáo dục, đánh giá chất lượng dịch vụ đào tạo trở lên thuận lợi. Mặt khác, nhằm nâng cao hiệu quả trong giảng dạy và học tập, đặc biệt là việc giảng dạy trực tuyến. Đây cũng là tiền đề hướng đến một hệ sinh thái thông minh kết hợp giữa công nghệ với các dịch vụ đào tạo dựa trên nền tảng số, trong đó bài toán xây dựng đại học thông minh là một giải pháp khả thi được các cơ sở đào tạo định hướng xây dựng. Để thực hiện được các công việc đó cần có một hạ tầng công nghệ thông tin đủ mạnh, đủ thông minh và đủ độ tin cậy nhằm giải quyết các bài toán, trong đó có việc xây dựng hệ thống tính toán hiệu năng cao HPC (High Performance Computing) nhằm tạo ra nền móng để triển khai các bài toán có nhu cầu xử lý tính toán lớn. Tuy nhiên, tại một số cơ sở đào tạo ở Việt Nam, việc đầu tư kinh phí xây dựng và duy trì một hệ thống như vậy gặp nhiều khó khăn về kinh phí, con người, hạ tầng... Để giải quyết vấn đề đó, bài báo này trình bày giải pháp xây dựng một hệ thống tính toán hiệu năng cao dựa trên các tài nguyên sẵn có sử dụng công nghệ ảo hóa Docker nhằm huy động sức mạnh tính toán từ các nền tảng phần cứng chưa sử dụng hết gọi tắt là hệ thống vHPC. Đề xuất này làm cơ sở khi triển khai các bài toán trong đại học thông minh.

* Tác giả liên hệ.

Email: truong@ueh.edu.vn (Trương Việt Phương), hungngq@ueh.edu.vn (Nguyễn Quốc Hùng), thanhln@ueh.edu.vn (Lê Ngọc Thanh), vhqđinh@ueh.edu.vn (Võ Hà Quang Định).

Trích dẫn bài báo: Trương Việt Phương, Nguyễn Quốc Hùng, Lê Ngọc Thanh, & Võ Hà Quang Định. (2020). Nghiên cứu xây dựng hệ thống tính toán hiệu năng cao phục vụ triển khai các bài toán trong Đại học thông minh. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 31(11), 53–65.

Keywords:

High Performance Computing;
Cluster Computing;
Docker;
Smart University;
vHPC

Abstract

Digital transformations in education is a topic attracting a lot of attention from many scientists and researchers in recent times. This transformation helps the computerization of the entire process of education management and assessment of the quality of training services become more favorable. Additionally, digital transformations also improve effectiveness in teaching and learning, especially in E-learning. This is also a premise towards an intelligent ecosystem combining technology with digital-based training services, in which the problem of building smart university is a possible solution. To perform these tasks, it is necessary to have an information technology infrastructure that is strong enough, intelligent enough, and reliable enough to solve problems in the overall framework, in which the construction of a computational system. High Performance Computing will aim to create a foundation for deploying problems with large computational processing needs. However, in some universities in Vietnam, investment in construction and maintenance of such a system faces many difficulties in terms of funding, people... This paper presents a solution to build a high-performance computing system based on available resources using Docker virtualization technology to mobilize computing power from underutilized hardware platforms, defined as a vHPC system. Therefore, this paper proposes an overall perspective when implementing problems smart university.

1. Đặt vấn đề

Ngày nay, việc sử dụng máy tính hiệu năng cao trong khoa học và kỹ thuật đã và đang thay đổi cơ bản hoạt động nghiên cứu khoa học. Nhiều ngành khoa học sử dụng tính toán như: Sinh học tính toán, hóa học tính toán, vật lý tính toán, vật liệu tính toán, cơ học tính toán, địa vật lý tính toán, thống kê tính toán, công nghệ tài chính ngân hàng... đã hình thành và phát triển nhanh chóng. Điểm chung nổi bật là các xử lý thông tin, phân tích và dự báo kết quả bằng tính toán, mô phỏng sử dụng hệ thống tính toán hiệu năng cao. Hệ thống này được xem là giải pháp quan trọng đáp ứng nhu cầu ngày càng cao về tính toán. Không chỉ trong nghiên cứu khoa học, rất nhiều ứng dụng thực tiễn trong nhiều lĩnh vực, đặc biệt là các lĩnh vực kinh tế - xã hội, đã và đang phát triển với quy mô lớn và công nghệ phức tạp hơn mà việc thực thi chúng trên các máy tính thông thường là không khả thi bởi khối lượng tính toán lớn và ràng buộc về thời gian.

Hệ thống tính toán hiệu năng cao HPC ra đời nhằm đáp ứng tình hình trên, ở đó cho phép thực thi các giải pháp dữ liệu lớn (Big Data) dựa trên hệ thống tích hợp nhiều nút tính toán trang bị chip xử lý CPU (Central Processing Unit) kết hợp với chip xử lý đồ họa GPU (Graphics Processing Unit). Một khía cạnh khác, khai thác tối đa nguồn tài nguyên trên các hệ thống tính toán lớn, HPC vẫn là bài toán cần được giải quyết trong một vài năm gần đây. Việc đi tìm lời giải tối ưu cho bài toán này nhằm khai thác tối đa khả năng tính toán của hệ thống HPC đang thu hút sự quan tâm của các nhà khoa học, các viện nghiên cứu, các trường đại học trong và ngoài nước. Tuy nhiên, để trang bị một hệ thống HPC cho một trường đại học hay một đơn vị nghiên cứu công lập đòi hỏi phải đầu tư một nguồn kinh phí

lớn để mua sắm, duy trì hệ thống hoạt động và đặc biệt cần có đội ngũ kỹ thuật có trình độ quản trị vận hành khi đưa vào sử dụng. Mặt khác, trên hệ thống vẫn tồn tại các thể hệ máy chủ đang hoạt động đơn lẻ, cài đặt các ứng dụng chạy độc lập. Do vậy, để xây dựng được hệ thống tính toán hiệu năng cao HPC thì một công nghệ ảo hóa các tài nguyên phần cứng sẵn có vHPC (Virtual High Performance Computing) được đề xuất.

Trong bài báo này, nhóm tác giả trình bày chi tiết công nghệ ảo hóa phần cứng sẵn có của các máy chủ đơn lẻ nhằm tạo ra một hệ thống tính toán hiệu năng cao sử dụng công nghệ Docker xử lý dữ liệu theo tính toán cụm máy tính gọi tắt là hệ thống vHPC.

2. Tổng quan nghiên cứu trong và ngoài nước

Tại Việt Nam, trong một vài năm gần đây, một số nghiên cứu trong lĩnh vực tính toán lớn hiệu năng cao đã được thực hiện như: Đề tài “*Nghiên cứu thiết kế hệ thống tính toán hiệu năng cao 50-100 TFLOPS*” (Thoại Nam, 2016) với mục tiêu xây dựng các giải pháp, các kỹ thuật được dùng trong lĩnh vực tính toán hiệu năng cao và tìm cách xây dựng một công cụ có thể đánh giá được hiệu năng của một hệ thống máy tính mạnh, hướng tới hiệu suất đạt khoảng 30% - 90% kết quả được công bố trên 500 máy tính mạnh nhất thế giới. Đề tài “*Nghiên cứu các hệ thống tính toán hiệu năng cao và ứng dụng mô phỏng vật liệu vi mô*” (Nguyễn Thanh Thủy, 2006) tại trung tâm tính toán hiệu năng cao trực thuộc Trường Đại học Bách khoa Hà Nội nằm trong khuôn khổ hợp tác khoa học công nghệ theo đề tài Nghị định Thư với Ấn độ giai đoạn 2004–2005 đã đề xuất xây dựng hệ thống tính toán cao gồm chức năng Bkluster có nhiệm vụ tính toán song song ghép cụm dựa trên kiến trúc Beowulf¹ và mô hình lập trình truyền thông điệp cùng với bộ phần mềm BKlusware là một tập các phần mềm hỗ trợ tối đa nhiều người dùng ở nhiều mức độ khác nhau.

Trong lĩnh vực ảo hóa hệ thống tính toán đã có một số đề tài/ dự án được đề xuất thực hiện như: Dự án “*Ứng dụng công nghệ ảo hóa và điện toán đám mây để xây dựng hệ thống lưu trữ và truyền hình ảnh (PACS-Cloud) phục vụ kết nối liên thông dữ liệu giữa các bệnh viện*” (Nguyễn Chí Ngọc, 2020) được Bộ Công Thương phê duyệt trong khuôn khổ Chương trình quốc gia phát triển một số ngành công nghiệp công nghệ cao. Trong nhiệm vụ khoa học mã số TNMT.2012.08.09, “*Nghiên cứu hệ thống hạ tầng ảo hóa và công nghệ điện toán đám mây riêng, cung cấp hạ tầng công nghệ thông tin trong ngành tài nguyên và môi trường*” của Trần Văn Đoàn (2012) thuộc Bộ Tài Nguyên Môi trường đã được đề xuất thực hiện các nhiệm vụ trọng tâm gồm: (1) Có được cơ sở khoa học của giải pháp điện toán đám mây và ứng dụng ảo hóa hạ tầng kỹ thuật thông tin do Cục Công nghệ thông tin đang quản lý; (2) ứng dụng được ảo hóa đối với server, thiết bị mạng và lưu trữ, thư viện điện tử ngành tài nguyên môi trường lên đám mây công cộng; (3) đề xuất định hướng đầu tư hạ tầng kỹ thuật, thông tin ngành tài nguyên và môi trường đến năm 2015 và định hướng đến năm 2020 với các sản phẩm như: Giải pháp ảo hóa và điện toán đám mây các hệ thống máy chủ, mạng, thiết bị lưu trữ tại Cục Công nghệ thông tin; Phần mềm ứng dụng giải pháp điện toán đám mây để đưa hệ thống thư viện điện tử ngành Tài nguyên và Môi trường đến độc giả trên đám mây công cộng; Đề xuất định hướng đầu tư hạ tầng ảo hóa và ứng dụng điện toán đám mây trong ngành Tài nguyên và Môi trường. Ngoài

¹ Beowulf là một cụm máy tính trong đó các máy tính phổ thông thường như nhau, nối mạng thành một mạng cục bộ nhỏ với các thư viện và các chương trình cài đặt cho phép chia sẻ quá trình xử lý giữa chúng. Kết quả có được là một cụm máy tính song song hiệu suất cao từ những máy tính cá nhân phổ thông.

ra, công nghệ ảo hóa được tổng hợp trong nghiên cứu “*Ứng dụng công nghệ ảo hóa trong việc tối ưu hóa cơ sở hạ tầng công nghệ thông tin*” của Hồ Văn Ngọc (2017) đã trình bày một số ứng dụng về công nghệ ảo hóa như ảo hóa Server, ảo hóa hạ tầng mạng, ảo hóa desktop, ảo hóa các ứng dụng, ảo hóa hệ thống lưu trữ, công nghệ ảo hóa điện toán đám mây... Từ đó đưa ra giải pháp và mô hình ứng dụng phù hợp cơ sở hạ tầng công nghệ thông tin dựa trên công nghệ ảo hóa (Vmwave Vsphere)...

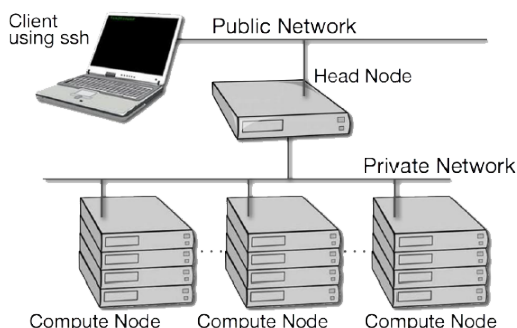
Trên thế giới, một số công nghệ ảo hóa hệ thống tính toán được áp dụng ở nhiều nơi trên thế giới như nghiên cứu của Bhanage và cộng sự (2011) về việc đánh giá thử nghiệm công nghệ ảo hóa OpenVZ (Open Virtuozzo) từ góc độ triển khai nhận xét rằng OpenVZ là một hệ thống công nghệ ảo hóa hoạt động dựa trên nhân Linux. OpenVZ cho phép một máy chủ vật lý để chạy nhiều trường hợp hệ điều hành riêng biệt, được gọi là container, máy chủ riêng ảo (VPSS), hoặc môi trường ảo (VES). OpenVZ không thực sự ảo hóa, nó sử dụng chung 1 nhân Linux đã được sửa đổi và do đó chỉ có thể chạy duy nhất hệ điều hành Linux. Như vậy tất cả các máy chủ ảo VPS cũng chỉ có thể chạy được Linux với chung một công nghệ và phiên bản Kernel. Tuy nhiên, do không có nhân riêng nên nó rất nhanh và hiệu quả, nhưng đó cũng chính là nhược điểm của nó khi tất cả các máy chủ phải sử dụng chung 1 nhân duy nhất. Công nghệ ảo hóa Hyper-V của hãng Microsoft (Finn, 2010) gồm 3 thành phần chính: Hypervisor, ngăn ảo hóa và mô hình I/O (nhập/xuất) ảo hóa mới. Hypervisor là lớp phần mềm rất nhỏ hiện diện ngay trên bộ xử lý (BXL) theo công nghệ Intel-V hay AMD-V, có vai trò tạo các phân vùng (Partition) mà thực thể ảo sẽ chạy trong đó. Nghiên cứu của Habib (2008) trình bày tổng quan về công nghệ ảo hóa KVM (viết tắt của Kernel Virtualization Machine) đề xuất công nghệ ảo hóa mới cho phép ảo hóa thực sự trên nền tảng phần cứng. Nó hoạt động tương tự như một người quản lý chia sẻ công bằng các tài nguyên như ổ đĩa (Disk), network IO và CPU. Máy chủ cài đặt Linux, nhưng KVM hỗ trợ tạo máy chủ ảo có thể chạy cả Linux, Windows. Nó cũng hỗ trợ cả hệ thống x86 và x86-64. Trong nghiên cứu của David (2007) đã trình bày công nghệ ảo hóa XEN là công nghệ ảo hóa thực sự cho phép chạy cùng lúc nhiều máy chủ ảo (VPS) trên 1 máy chủ vật lý. Mỗi VPS có hệ thống File System riêng và hoạt động như 1 máy chủ vật lý độc lập. Tiếp đó là công nghệ ảo hóa tối ưu và được sử dụng nhiều nhất trong thời gian gần đây, công nghệ ảo hóa VMWare (Jin, 2009) do công ty VMWare phát triển, nó hỗ trợ ảo hóa từ mức phần cứng. Giao diện thân thiện, cách cài đặt và sử dụng đơn giản, nhiều tính năng cao cấp, hỗ trợ nhiều hệ điều hành, các phiên bản đa dạng. Công nghệ này thường áp dụng cho các đơn vị lớn như ngân hàng và ít được sử dụng cho các VPS thương mại bán trên thị trường. Cấu trúc của nó là một chương trình ứng dụng ảo hóa, chạy trên nền hệ điều hành Linux hay Windows. Trong nghiên cứu của Yu và Huang (2015) đã đề xuất giải pháp tự động xây dựng cầu kết nối các nút tính toán (Computer Node) thành các cụm có khả năng xử lý song song (Message Passing Interface – MPI) trong hệ thống HPC sử dụng Docker Cluster, công việc này cũng được triển khai bởi De Bayser và Cerqueira (2017) nhằm tích hợp tính toán song song MPI với hệ thống HPC sử dụng Docker. Việc đánh giá hiệu quả khi đề xuất ảo hóa hệ thống tính toán làm căn cứ xây dựng HPC là vấn đề quan trọng, Abdullah và cộng sự (2019) đã tiến hành nghiên cứu đánh giá tính hiệu quả khi sử dụng máy ảo (Virtual Machines – VMs)² và Docker trên nền dịch vụ điện toán đám mây OpenNebula. Kết quả cho thấy khi triển khai ảo hóa bằng Docker hiệu năng đạt 70,23% so với VMs là 46,48%, từ đó chứng minh được hiệu quả khi ảo hóa dùng Docker.

² Máy ảo (Virtual Machine – VMs) là một chương trình giả lập một hệ thống máy tính, được chạy trên hệ điều hành chủ và hoạt động như một máy tính vật lý. Một máy ảo cung cấp phần cứng ảo (bao gồm bộ xử lý trung tâm CPU, bộ nhớ RAM, ổ đĩa cứng HDD) để chạy hệ điều hành và các phần mềm riêng trên đó. Các phần cứng ảo này được ánh xạ tới phần cứng thực trên máy tính vật lý và chia sẻ tài nguyên phần cứng với máy vật lý.

3. Hệ thống tính toán hiệu năng cao và công nghệ ảo hóa

3.1. Giới thiệu hệ thống tính toán hiệu năng cao

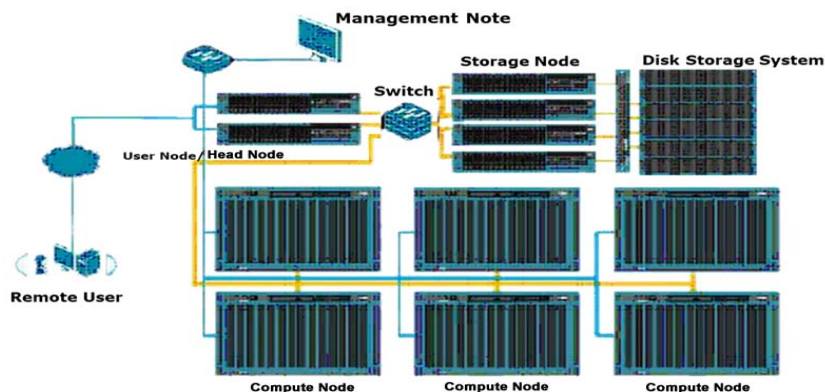
HPC là một hệ thống tính toán hiệu năng cao được tích hợp nhiều máy tính lại với nhau nhằm giải quyết các bài toán lớn có độ phức tạp tính toán cao, minh họa như hình dưới đây:



Hình 1. Mô hình tổng thể của hệ thống HPC

Nguồn: Nwogbaga và cộng sự (2016)

Trong Hình 1 các máy tính được kết nối với nhau qua mạng nội bộ (Private Network) và kết nối với một máy tính điều khiển (Head Node) hay gọi là nút điều khiển, nút này kết nối ra ngoài bằng mạng công khai (Public Network) qua giao thức bảo mật SSH. Mỗi máy tính tham gia trong hệ thống HPC được định nghĩa là một nút tính toán, một hệ thống HPC gồm nhiều nút tính toán và có thể bổ sung thêm nhiều nút tính toán nếu cần thiết nhằm nâng cao hiệu năng hệ thống.



Hình 2. Kiến trúc của hệ thống HPC triển khai

Nguồn: Li (2015)

Kiến trúc của hệ thống HPC trong Hình 2 được chia ra làm 02 cách xử lý:

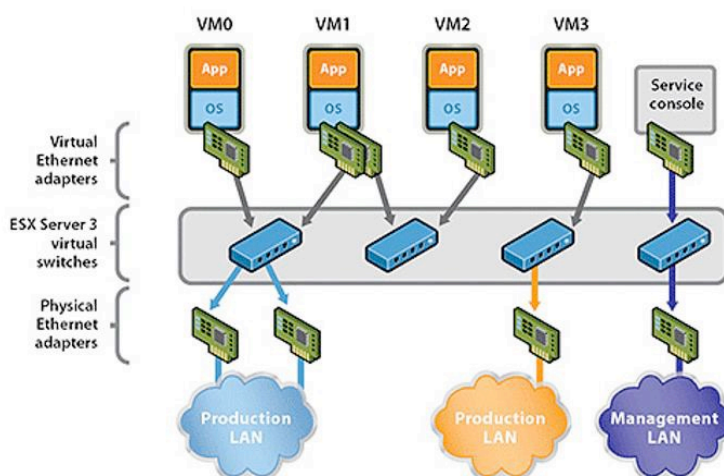
- Xử lý song song MPP (Massive Parallel Processing): Phân bố các công việc xử lý trên nhiều bộ vi xử lý trên một Card đồ họa GPU chuyên dụng. Với hàng (trăm) ngàn vi xử lý kết nối qua đường mạng tốc độ cao.

- Xử lý cụm máy tính CC (Cluster Computing): Gồm nhiều server riêng lẻ được liên kết và hoạt động cùng với nhau trong một hệ thống, dựa trên tính toán các cụm máy tính chủ yếu nhờ vào bộ xử lý trung tâm CPU.

3.2. Công nghệ ảo hóa

Ảo hóa là quá trình tạo ra một bản sao ảo của một thực thể nào đó, được thiết kế để tạo ra tầng trung gian giữa hệ thống phần cứng và phần mềm; giúp quản lý, phân phối tài nguyên phần cứng cho lớp ứng dụng mà công nghệ ảo hóa mang lại, gồm có các loại ảo hóa sau:

Ảo hóa mạng (Network) trong nghiên cứu của Lee (2014) là một tiến trình hợp nhất tài nguyên, thiết bị mạng cả phần cứng lẫn phần mềm thành một hệ thống mạng ảo. Sau đó, các mạng ảo này sẽ tạo ra một vùng làm việc độc lập mà chỉ có các địa chỉ được định danh vào vùng này như minh họa trong hình dưới đây:

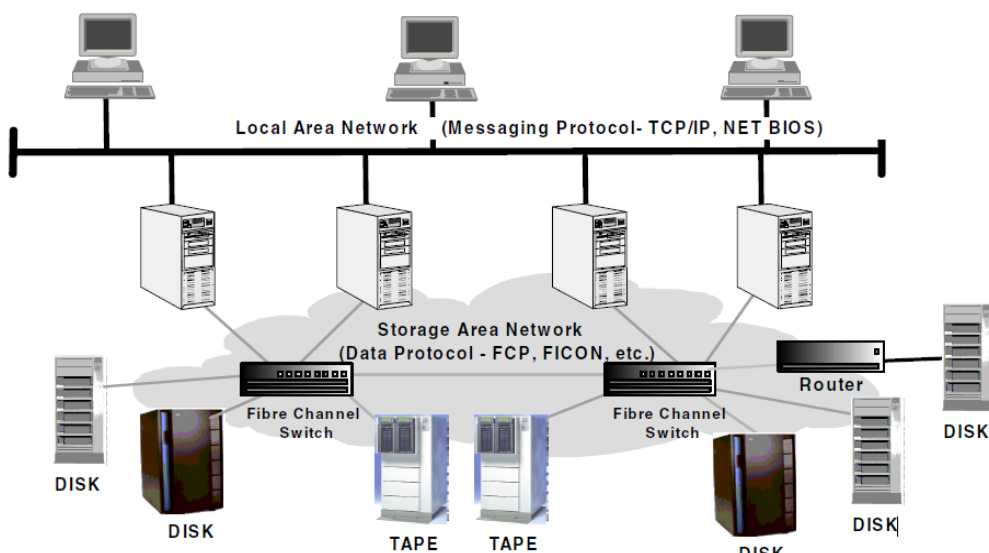


Hình 3. Ảo hóa mô hình mạng (Network)

Nguồn: Lee (2014)

Một máy ảo có thể có nhiều Card mạng ảo, việc tạo các Card mạng ảo không giới hạn, có thể nối các máy ảo này lại với nhau bằng một Switch ảo. Tốc độ truyền dữ liệu giữa các máy ảo với nhau thông qua các Switch ảo với tốc độ rất cao theo chuẩn GIGABITE, dẫn đến việc đồng bộ giữa các máy ảo với nhau diễn ra rất nhanh.

- Ảo hóa lưu trữ (Storage) trong nghiên cứu (Khattar và cộng sự, 1999) về việc ảo hóa lưu trữ từ các thiết bị vật lý như băng từ, ổ cứng hay kết hợp cả 2 loại. Việc làm này mang lại các ích lợi như việc tăng tốc khả năng truy xuất dữ liệu, do việc phân chia các tác vụ đọc, viết trong mạng lưu trữ. Ngoài ra, việc mô phỏng các thiết bị lưu trữ vật lý cho phép tiết kiệm thời gian hơn thay vì phải định vị xem máy chủ nào hoạt động trên ổ cứng nào để truy xuất. Kết quả của nghiên cứu này đã đưa ra nhiều đóng góp mang tính chất ứng dụng trong ảo hóa lưu trữ mà công nghệ ảo hóa SAN (Storage Area Network) là một sản phẩm khá nổi bật trong thời gian qua và được minh họa như hình dưới đây:



Hình 4. Ảo hóa mô hình lưu trữ (Storage)

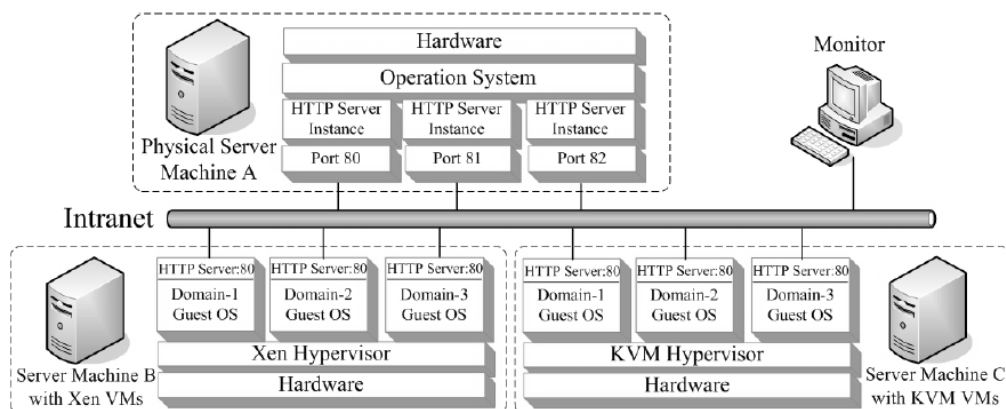
Nguồn: Khattar và cộng sự (1999)

Trong Hình 4, hệ thống ảo hóa lưu trữ SAN có hai đặc điểm cơ bản:

- Mạng (Network) có tác dụng truyền thông tin giữa thiết bị lưu trữ và hệ thống máy tính. Một SAN bao gồm: 1 cấu trúc truyền tin – nó cung cấp kết nối vật lý, và quản lý các lớp, tổ chức các kết nối, các thiết bị lưu trữ, và hệ thống máy tính sao cho dữ liệu truyền trên đó với tốc độ cao và bảo mật;

- Một hệ thống lưu trữ bao gồm các thiết bị lưu trữ, hệ thống máy tính, hay các ứng dụng chạy trên nó, và một phần rất quan trọng là các phần mềm điều khiển, quá trình truyền thông tin qua mạng.

- Ảo hóa máy chủ (Server) trong nghiên cứu của Jin và cộng sự (2012) là một phương pháp phân vùng một máy chủ vật lý thành nhiều máy chủ ảo, mỗi máy chủ có khả năng của riêng của mình chạy trên máy tính dành riêng. Mỗi máy chủ ảo riêng của nó có thể chạy hệ điều hành khác nhau, và mỗi máy chủ ảo có thể được khởi động lại độc lập như hình minh họa dưới đây:



Hình 5. Ảo hóa mô hình máy chủ (Server)

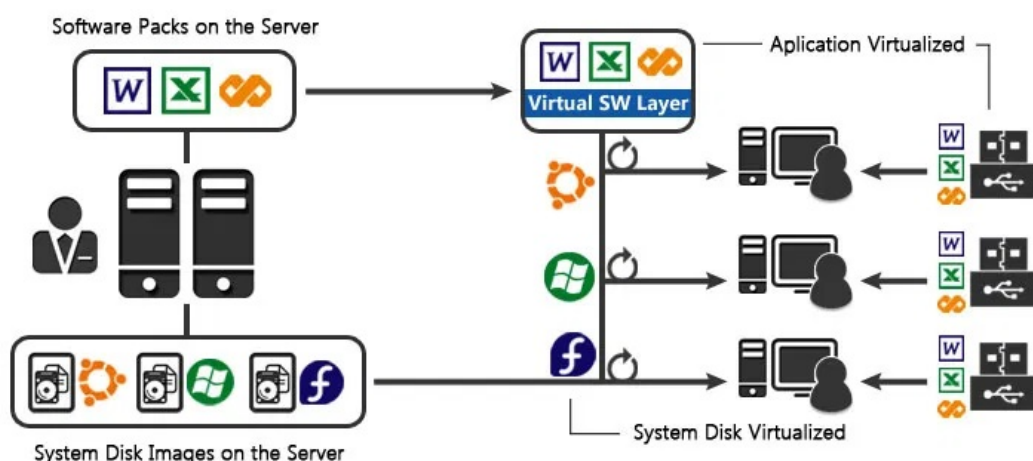
Nguồn: Jin và cộng sự (2012)

Trong Hình 5 mô tả hai môi trường máy chủ ảo hóa: Ảo hóa toàn phần (Full Virtualization), và ảo hóa một nửa (Paravirtualization).

- Ảo hóa toàn phần: Phần cứng được mô phỏng để mở rộng chạy những hệ điều hành khách trên nền tảng ảo hóa. Điều này có nghĩa rằng các thiết bị phần cứng khác nhau đều được mô phỏng. Thông thường, có nhiều nền tảng ảo hóa cố gắng chạy nhiều sự ủy nhiệm trên CPU chính (chạy nhanh hơn nhiều so với CPU mô phỏng) nhằm nắm bắt và xử lý các sự ủy nhiệm một cách thích hợp.

- Ảo hóa một nửa: Là phương pháp ảo hóa máy chủ khác, thay vì mô phỏng một môi trường phần cứng hoàn chỉnh, phần mềm ảo hóa này tạo ra một lớp Hypervisor³ nhằm mục đích dồn các truy cập các hệ điều hành máy chủ vào tài nguyên máy vật lý cơ sở.

- Ảo hóa ứng dụng (Application) trong nghiên cứu của Kusnetzky (2011) đã giới thiệu về nguyên tắc khi thực hiện ảo hóa ứng dụng, trong đó tách rời mối liên kết giữa ứng dụng và hệ điều hành, cho phép phân phối lại ứng dụng phù hợp với nhu cầu người dùng. Một ứng dụng được ảo hóa sẽ không được cài đặt lên máy tính một cách thông thường, mặc dù ở góc độ người dùng thì ứng dụng vẫn hoạt động một cách bình thường. Ứng dụng đó được cài đặt trên các máy chủ nhằm quản lý việc cập nhật phần mềm trở nên dễ dàng hơn, giải quyết sự đụng độ giữa các ứng dụng và việc thử nghiệm sự tương thích của chúng cũng trở nên dễ dàng hơn:



Hình 6. Mô hình ảo hóa ứng dụng

Nguồn: Kusnetzky (2011)

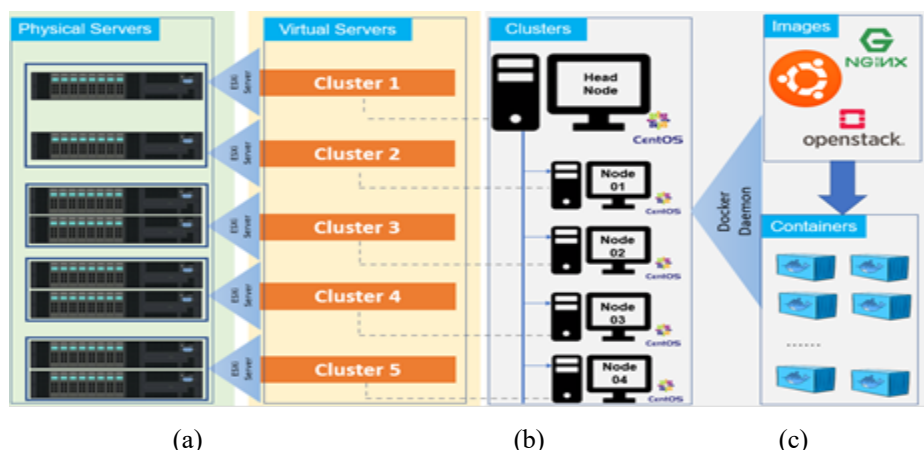
Hình 6 cho thấy ưu điểm của ảo hóa ứng dụng đó là tất cả các máy tính đều có thể sử dụng phần mềm ảo như đang cài trên máy tính của mình mà không phải lo về cấu hình; thông tin luôn luôn được lưu trữ an toàn ở máy chủ trung tâm thay vì có thể phân tán ra từng máy trạm; phân phối phần mềm một cách linh động đến một số cá nhân hoặc nhóm có nhu cầu sử dụng thay vì cài vào tất cả mọi máy như cách phổ thông; phân phối hoặc gỡ bỏ phần mềm cài đặt lên hệ thống một cách nhanh chóng.

³ Hypervisor là phần mềm giám sát máy ảo (VMs): Là một chương trình phần mềm quản lý một hoặc nhiều máy ảo. Các hypervisor cho phép mỗi VM truy cập vào lớp tài nguyên phần cứng vật lý như CPU, RAM và lưu trữ HDD. Nó cũng có thể giới hạn số lượng tài nguyên hệ thống mà mỗi máy ảo có thể sử dụng để đảm bảo cho nhiều máy ảo cùng sử dụng đồng thời trên một hệ thống.

4. Đề xuất ảo hóa hệ thống tính toán hiệu năng cao

4.1. Khung làm việc tổng quát

Trong nghiên cứu này, nhóm tác giả đề xuất ảo hóa phần cứng các máy chủ vật lý nhằm tạo ra các máy chủ ảo đóng vai trò như một cụm máy tính xử lý. Sau đó, sử dụng nền tảng quản lý Docker nhằm tạo ra hệ thống tính toán hiệu năng cao gọi tắt là vHPC (Virtual High-Performance Computing) được mô tả như hình dưới đây:



Hình 7. Khung làm việc hệ thống vHPC

Hình 7 minh họa kiến trúc hệ thống vHPC gồm có 3 lớp xử lý:

Lớp vật lý: Được tích hợp từ 8 server vật lý thương hiệu IBM, HP, DELL được minh họa trong Hình 7(a) và có cấu hình như Bảng 1 dưới đây:

Bảng 1.

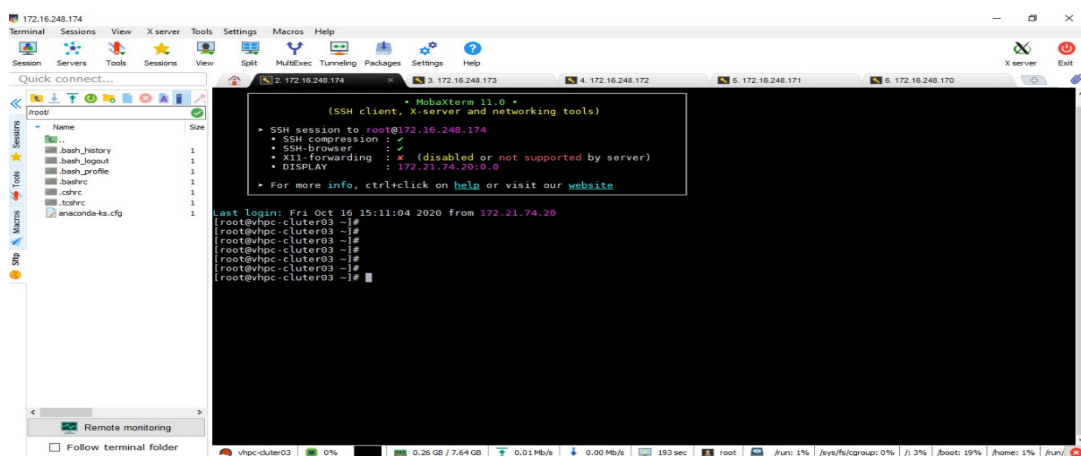
Cấu hình máy chủ vật lý

STT	Hãng sản xuất và thông số kỹ thuật	Khả năng lưu trữ
1	IBM X3560 M4, 12 Core 2.0GHz, E5 2620, 24G Ram	270GB/272GB
2	HP Proliant DL380 Gen9, 32 Core 2.10GHz, E5 2620, 64G Ram	600GB/830GB
3	HP Proliant DL380 Gen9, 24 Core 2.10GHz, E5 2620, 128G Ram	100GB/830GB
4	IBM X3560 M3, 16 Core 2.4GHz, E5620, 36G Ram	1.62TB/1.62TB
5	DELL PowerEdge R740, 24 Core 2.59GHz, 64G Ram	1.30TB/1.90TB
6	DELL PowerEdge R720, 24 Core 2.10GHz, 64G Ram	2TB/3.26TB
7	DELL PowerEdge R720, 24 Core 2.10GHz, 64G Ram	1TB/3.26TB
8	SAN (Storage Area Network)	7TB

Nguồn: Khảo sát thực tế tại UEH

Tuy nhiên, trong thực tế khi khảo sát cho thấy các máy chủ vật lý (Server) vẫn đang hoạt động và duy trì một số phần mềm chuyên ngành, vẫn còn một lượng tài nguyên chưa dùng đến.

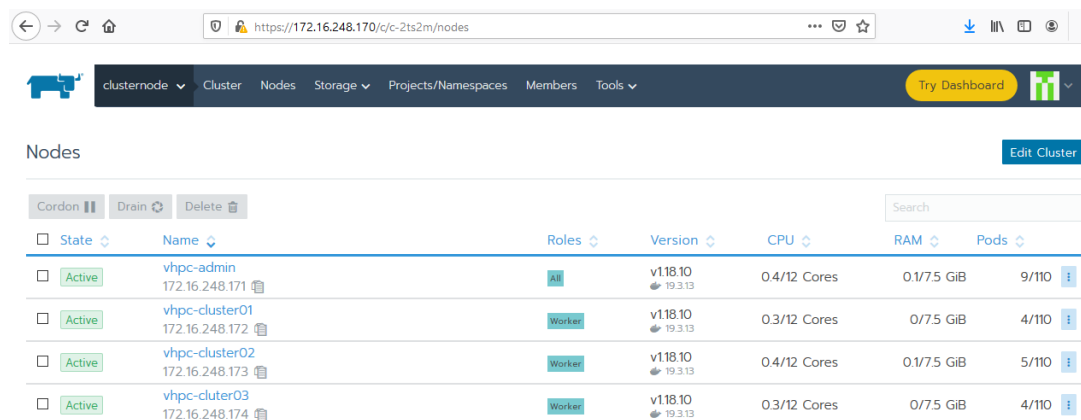
Lớp ảo hóa: Căn cứ vào báo cáo đánh giá phân tích của Taknet (Tan và cộng sự, 2013) về việc xây dựng một hệ thống HPC thì cấu hình tối thiểu bao gồm: 1 máy tính điều khiển (Head Node) và ít nhất từ 3 đến 4 máy tính xử lý (Compute Node), các nút này hoàn toàn có thể bổ sung nếu cần thiết phải nâng cao hiệu năng tính toán. Vì vậy, nhóm tác giả áp dụng công nghệ ảo hóa máy chủ tạo ra 05 máy chủ ảo được minh họa trong Hình 7(b), trong đó, 01 máy đóng vai trò là quản lý (Head Node) và 04 máy còn lại đóng vai trò là cụm máy tính xử lý (Compute Node). Tất cả các máy tính được cài đặt hệ điều hành CenOS v7.8 như hình dưới đây:



Hình 8. Ảo hóa các cụm máy tính xử lý trong hệ thống vHPC

Nguồn: Kết quả nghiên cứu hệ thống vHPC

Lớp tính toán: Nhóm tác giả lựa chọn công nghệ Docker để tiến hành xây dựng hệ thống vHPC như Hình 7(c) bao gồm hai thành phần: (1) Phần hệ điều hành CenOS v7.8 chạy nền ở phía dưới; (2) Giao diện tương tác nhằm cung cấp cho tài nguyên tính toán bao gồm các gói dữ liệu có sẵn (Image) và các kho chứa dữ liệu (Containers) minh họa như hình dưới đây:

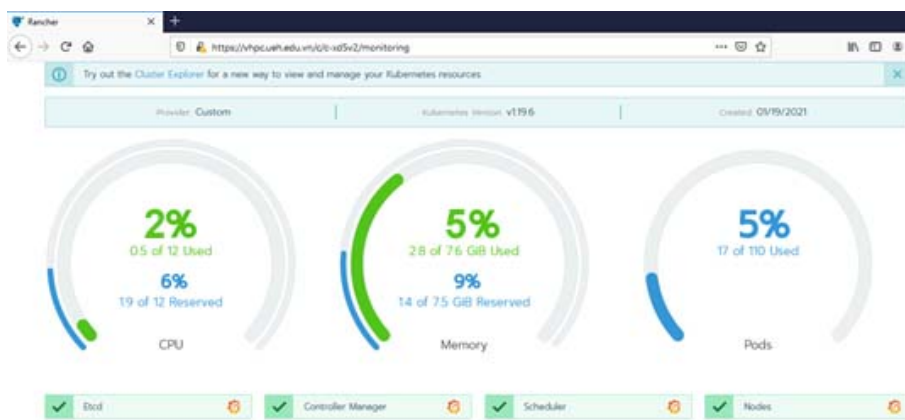


Hình 9. Hệ thống vHPC xây dựng trên nền tảng máy ảo sử dụng Docker

Nguồn: Kết quả nghiên cứu hệ thống vHPC

4.2. Đánh giá hiệu năng hệ thống vHPC

Hệ thống vHPC hoạt động trên nền tảng Webform tại địa chỉ <http://vHPC.ueh.edu.vn>, kết quả đánh giá về hiệu năng được thông kê như hình dưới đây:



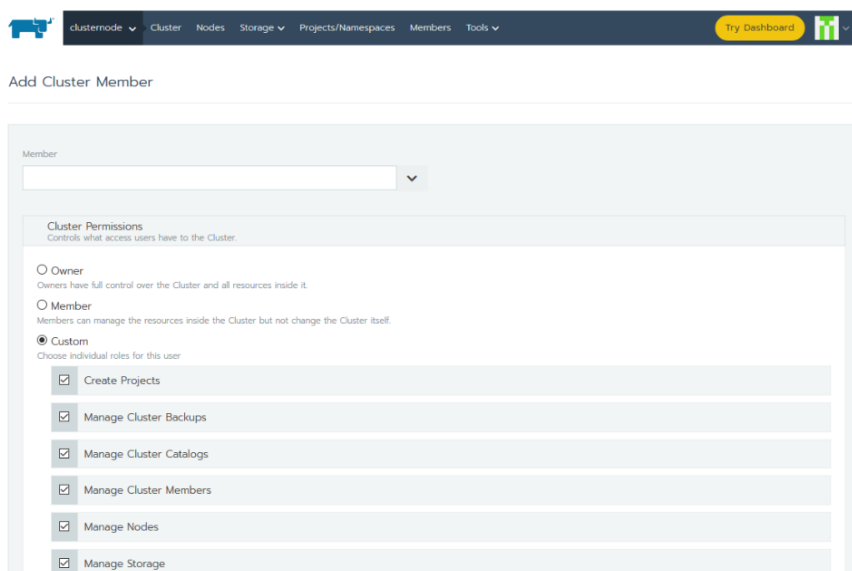
Hình 10. Hiệu năng hệ thống vHPC khi hoạt động

Nguồn: Kết quả nghiên cứu hệ thống vHPC

Trong Hình 10 hiệu năng của hệ thống chiếm 5% bộ nhớ, tài nguyên bộ nhớ sử dụng 2% và nhu cầu tính toán của người sử dụng vHPC đạt 5% khi ở chế độ tiêu chuẩn.

4.3. Tương tác giữa người dùng và hệ thống vHPC

Phần tương tác với người dùng, hệ thống vHPC cung cấp các tài khoản cho các thành viên được người quản trị thực hiện trên cùng giao diện khai thác, tương ứng với yêu cầu phần cứng tính toán được xác định quyền hạn của trong gói dữ liệu và cài đặt vào hệ thống, minh họa như Hình 11:



Hình 11. Cấp tài khoản thành viên sử dụng hệ thống vHPC

Nguồn: Kết quả nghiên cứu hệ thống vHPC

5. Kết luận

Công nghệ ảo hóa được phát triển và sử dụng trong khoảng thời gian khá lâu với nhiều ưu điểm. Ngoài sự linh động trong việc triển khai, máy ảo còn được biết đến với những đặc điểm như: Dễ dàng quản lý, khả năng bảo mật cao, hiệu suất về sự cô lập giữa phạm vi sử dụng cũng như quyền kiểm soát. Hiện nay, trong lĩnh vực tính toán hiệu năng cao, hạ tầng triển khai máy ảo đóng một vị trí quan trọng đối với các dịch vụ như điện toán đám mây, lưu trữ đám mây...

Trong bài báo này, nhóm tác giả đã đề xuất ảo hóa hệ thống vHPC dựa trên nguồn tài nguyên phần cứng sẵn có tại Trường Đại học Kinh tế TP. Hồ Chí Minh và tiến hành một số thử nghiệm công nghệ ảo hóa máy chủ tích hợp các bộ nhớ xử lý tính toán lớn như CPU và GPU nhằm hướng tới một hệ thống tính toán hiệu năng cao HPC với những tính năng ưu việt, khả năng mở rộng, và độ tin cậy cao hơn. Với đề xuất này, nhóm tác giả mong muốn giúp cho các nhà nghiên cứu, nhà quản lý, nhà khoa học có được các công cụ để thử nghiệm các giải thuật học máy trên nguồn dữ liệu lớn tính toán trên các môi trường khác nhau nhằm đáp ứng các yêu tố về mặt kỹ thuật như tính sẵn sàng cao, tính năng chịu lỗi ưu việt của hệ thống đề xuất■

Lời cảm ơn

Bài báo này được tài trợ bởi Trường Đại học Kinh tế TP. Hồ Chí Minh theo Đề tài NCKH cấp trường, mã số: CS-2020-14, tên đề tài: “Nghiên cứu và xây dựng hệ thống tính toán xử lý dữ liệu Big Data trên nền tảng Virtual High Performance Computing”. Chủ nhiệm đề tài: TS. Nguyễn Quốc Hùng.

Tài liệu tham khảo

- Abdullah, M., Iqbal, W., & Bukhari, F. (2019). Containers vs virtual machines for auto-scaling multi-tier applications under dynamically increasing workloads. In Bajwa, I., Kamareddine, F., & Costa, A. (Eds), *Intelligent Technologies and Applications. INTAP 2018. Communications in Computer and Information Science*, 932. Singapore: Springer. doi: 10.1007/978-981-13-6052-7_14
- Bhanage, G., Seskar, I., Zhang, Y., Raychaudhuri, D., & Jain, S. (2011). Experimental evaluation of OpenVZ from a testbed deployment perspective. In Magedanz, T., Gavras, A., Thanh, N. H., Chase, J. S. (eds). *Testbeds and Research Infrastructures. Development of Networks and Communities. TridentCom 2010. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, 46. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-17851-1_7
- David, C. (2007). *The definitive guide to the xen hypervisor (Prentice Hall open source software development series)*. US: Prentice Hall.
- De Bayser, M., & Cerqueira, R. (2017). *Integrating MPI with Docker for HPC*. Paper presented at the 2017 IEEE International Conference on Cloud Engineering (IC2E). doi: 10.1109/IC2E.2017.40
- Finn, A. (2010). *Mastering Hyper-V Deployment*. US: John Wiley & Sons.

- Habib, I. (2008). Virtualization with KVM. *Linux Journal*, 166(8).
- Hồ Văn Ngọc. (2017). Ứng dụng công nghệ ảo hóa trong việc tối ưu hóa cơ sở hạ tầng công nghệ thông tin. *Tạp chí Khoa học trường Đại học Văn Hiến*, 5(5), 112–118.
- Jin, S. (2009). *VMware VI and vSphere SDK: Managing the VMware infrastructure and vSphere*. US: Pearson Education.
- Jin, Y., Wen, Y., & Chen, Q. (2012). *Energy Efficiency and Server Virtualization in Data Centers: An Empirical Investigation*. Paper presented at the 2012 Proceedings IEEE INFOCOM Workshops. doi: 10.1109/INFCOMW.2012.6193474
- Khattar, R. K., Murphy, M. S., Tarella, G. J., & Nystrom, K. E. (1999). *Introduction to Storage Area Network, SAN*. California: IBM Corporation, International Technical Support Organization.
- Kusnetzky, D. (2011). *Virtualization: A Manager's Guide*. CA: O'Reilly Media Inc.
- Lee, H. (2014). Virtualization basics: Understanding techniques and fundamentals. In *School of Informatics and Computing Indiana University*. Retrieved from <http://cgl.soic.iu.edu/publications/virtualization.pdf>
- Thoại Nam. (2017). *Nghiên cứu thiết kế hệ thống tính toán hiệu năng cao 50-100 TFlops*. Đề tài nghiên cứu khoa học cấp trường, Trường Đại học Bách khoa, Đại học Quốc gia TP. Hồ Chí Minh.
- Nguyễn Chí Ngọc. (2020). *Ứng dụng công nghệ ảo hóa và điện toán đám mây để xây dựng hệ thống lưu trữ và truyền hình ảnh (PACS-Cloud) phục vụ kết nối liên thông dữ liệu giữa các bệnh viện*. Đề tài nghiên cứu khoa học cấp Bộ, Bộ Công thương.
- Nguyễn Thanh Thủy. (2006). *Nghiên cứu các hệ thống tính toán hiệu năng cao và ứng dụng mô phỏng vật liệu vi mô*. Trong khuôn khổ hợp tác Khoa học Công nghệ theo Nghị định Thư với Ấn Độ giai đoạn 2004–2005.
- Tan, G., Yeo, G. K., Turner, S. J., & Teo, Y. M. (Eds.). (2013). *AsiaSim 2013*. In *13th International Conference on Systems Simulation, Singapore, November 6–8, 2013*. Singapore: Springer.
- Trần Văn Đoàn. (2012). *Nghiên cứu hệ thống hạ tầng ảo hóa và công nghệ điện toán đám mây riêng, cung cấp hạ tầng công nghệ thông tin trong ngành tài nguyên và môi trường*. Đề tài nghiên cứu công nghệ thông tin cấp Bộ, Cục Công nghệ thông tin và dữ liệu Tài nguyên Môi trường.
- Yu, H.-E., & Huang, W. (2015). *Building a virtual HPC cluster with auto scaling by the docker*. Cornell University. Retrieved from <https://arxiv.org/abs/1509.08231>
- Nwogbaga, N., Mbanefo, E., & Nwoyibe, O. (2016). Critical analysis of cloud computing and its advantages over other computing techniques. *Journal Computing*, 3(2), 3955–3960.
- Li, L. (2015). *High Performance Computing Applied to Cloud Computing*. Thesis Technology, Communication and Transport Programme in Information Technology, Lapland University of Applied Sciences.