ETRI

2013
Electronics and
Telecommunications
Trends

빅데이터 처리 및
분석 기술 특집

차세대 클라우드 컴퓨팅을 위한 패브릭 컴퓨팅 기술

Fabric Computing for Next Generation Cloud Computing

이중수 (J. Lee) BigData 시스템구조연구팀 선임연구원
안창원 (C.W. Ahn) BigData 시스템구조연구팀 팀장

* 본 연구는 지식경제부 한국산업기술진흥원 유럽다자간 국제공동기술개발사업의 지원을 받아 수행 중인 '고성능 임베디드 시스템을 위한 매니코아 프로그래밍 자원 관리 기술 개발'(ITEA2 10021) 과제의 연구 결과임.

클라우드 컴퓨팅의 구현과 관리의 차세대 모델로 지목되는 패브릭 컴퓨팅은 오랜 컴퓨터 전문가들의 바람에도 불구하고 아직 구현 단계에 접근하지 못하였다. 그러나, 최근 패브릭 컴퓨팅에 대한 관심이 높아지고 이를 위한 초보적인 시도가 이루어지고 있다. 본고에서는 패브릭 컴퓨팅의 개념과 필요한 요소 기술, 관련 제품을 살펴보고 기술 발전 방향을 가늠하고자 한다.

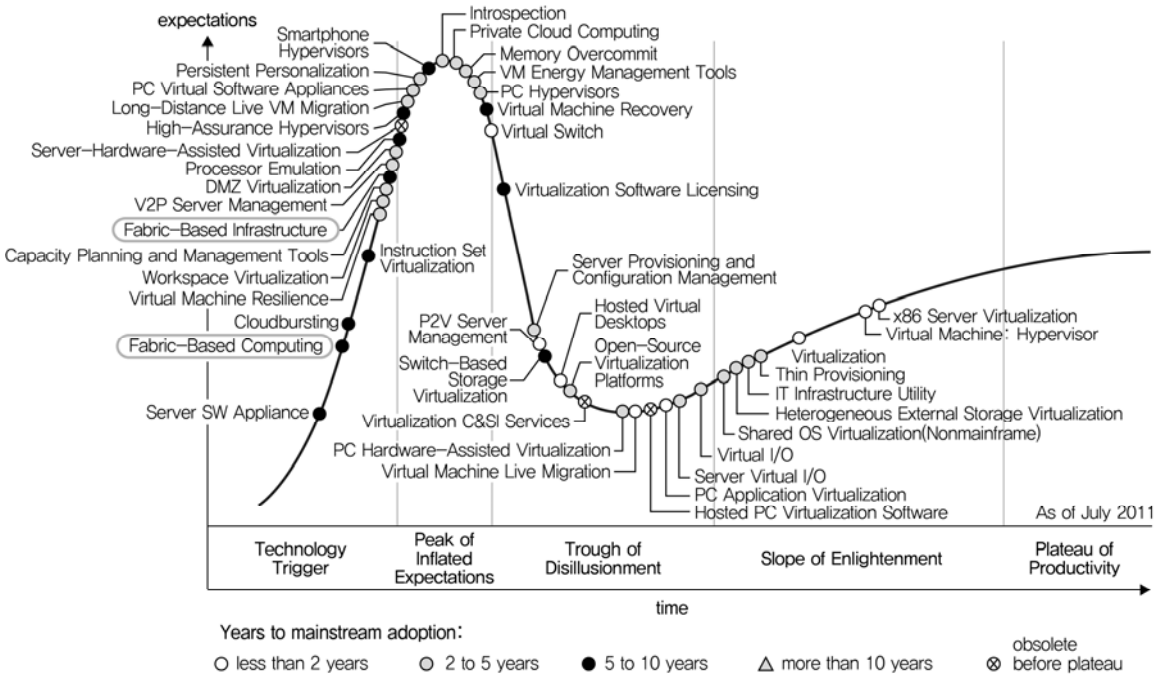
Ⅰ. 서론

클라우드 컴퓨팅을 활용한 비즈니스 애플리케이션이 다양해지고 산업에서의 활용도가 높아짐에 따라 컴퓨팅 자원을 효율적으로 유지하고 관리하고자 하는 요구가 높아지고 있다. 워크로드에 따라 동적으로 자원을 할당하고 구성/재구성하는 것이 클라우드 컴퓨팅의 장점이지만, 수용하기 어려운 응용 분야가 존재하며 자원 단위가 아니라 서버 노드 단위의 확장을 통해 컴퓨팅 용량의 증설을 제공하는 실정이다.

패브릭 컴퓨팅은 하나의 관리 체계로 전체 클라우드 컴퓨팅 인프라를 관리하고 구성 자원(프로세서, 메모리, I/O) 단위로 확장하거나 관리하고자 하는 요구에 의해 출발하였다. 하나의 컴퓨터를 여러 개의 가상머신으로 분할하여 제공하기보다 필요한 만큼의 자원을 충분히 제공할 수 있도록 물리 자원의 조직화와 동적 구성을 지원하게 된다. 서버 기능과 진화된 네트워킹을 결합하여 사용자에게 제공하는 엔터프라이즈 서버의 차세대 구조가 패브릭 컴퓨팅 기술이다[1].

가트너(Gartner)에서는 2011년에 패브릭 컴퓨팅을 10대 기술로 선정하여 데이터센터를 위한 새로운 컴퓨팅 패러다임의 출현을 예고하였다. (그림 1)에 도시된 2011년 "Hype-cycle"에 따르면 '패브릭 기반 컴퓨터(fabric-based computer)'와 '패브릭 기반 인프라(fabric-based infrastructure)'는 기술 발전 단계에서 초기에 해당한다.

본고에서는 클라우드 컴퓨팅 패러다임의 진화 방향으로서의 패브릭 컴퓨팅 기술에 대한 관련 기술과 최근 동향에 대해 설명하고자 한다. Ⅱ장에서는 패브릭 컴퓨팅의 두 가지 개념(패브릭 기반 컴퓨터와 패브릭 기반 인프라)에 대해서 설명하고 Ⅲ장에서는 패브릭 컴퓨팅을 실현하기 위해 필요한 요소 기술에 대해 논의한다. Ⅳ장에서는 패브릭 기반 인프라를 위해 등장한 몇 가지 제품의 사례를 살펴 본다. 마지막으로, Ⅴ장에서 패브릭 컴퓨팅의 한계와 전망으로 결론을 맺는다.



〈자료〉: Gartner, 2011.

(그림 1) 패브릭 컴퓨팅 Hype-cycle

II. 패브릭 컴퓨팅의 개념

패브릭 컴퓨팅은 컴퓨팅 시스템을 구성하는 자원(프로세서, 메모리, I/O)을 필요한 만큼, 독립적으로, 즉시 구성하여 사용하고자 하는 요구에서 도출된 개념이다. 클라우드 컴퓨팅에서의 사례와 같이 컴퓨터 시스템의 병렬화와 모듈화가 지속됨에 따라 패브릭화가 가속되고 있으며 궁극적으로는 패브릭 컴퓨팅에 의해 컴퓨팅 시스템이 즉시 조립되어 제공되는 방향으로 진화할 것이라는 예상이 가능하다[2].

'패브릭(fabric)'이라는 용어가 사용되는 분야가 다양하여 혼동의 여지가 많지만 패브릭 컴퓨팅의 개념은 다수의 노드(프로세서, 메모리, I/O 장치 등)와 이를 연결하는 링크가 있는 시스템을 표현하기 위해 사용한다.

위키피디아에서는 '패브릭 컴퓨팅'이 다음과 같이 정의되어 있다.

"느슨히 결합된 저장 장치, 네트워킹, 프로세싱 기능이 높은 대역폭의 연결망으로 연결되어 있는 고성능 컴퓨팅 시스템"[3]

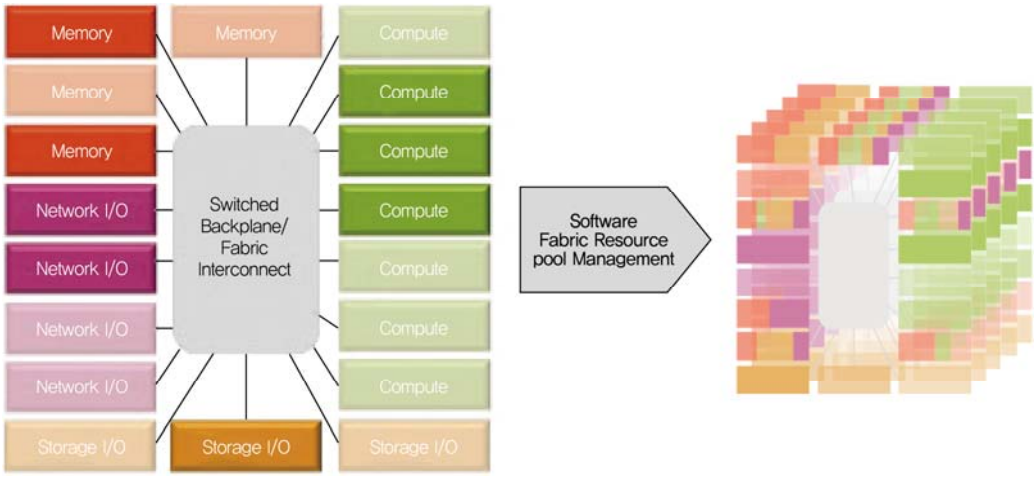
패브릭 컴퓨팅에 대한 전망과 연구보고서를 활발히 내놓고 있는 가트너에서는 다음과 같이 세분화된 패브릭 컴퓨팅에 관한 정의를 내리고 있다[4].

패브릭 기반 컴퓨터(Fabric-Based Computer: FBC)는 패브릭 인터커넥트 혹은 백플레인 스위치에 연결되어 있는 빌딩 블록 모듈을 조합하여 제공할 수 있는 모듈화된 형태의 컴퓨팅 시스템이다.

패브릭 가능 컴퓨터(Fabric-Enabled Computer: FEC)는 패브릭 기반 컴퓨터(FBC)로 넘어가는 중간 단계의 컴퓨터 형태이다. 프로세싱 블록, 저장 장치, 네트워킹, 관리 소프트웨어의 부분적인 통합을 제공하며 현재의 블레이드 서버가 이에 해당한다.

패브릭 가능 컴퓨터는 정적인 경우와 동적인 경우로 나누어 생각할 수 있는데, 정적인 경우는 하드웨어의 구조가 현재와 유사하나 소프트웨어에 의해 자원의 배분이 이루어지는 Google의 skinless x86 서버를 예로 들 수 있다. 동적인 경우는 하드웨어 수준의 분할과 통합이 이루어지는 경우로 아직 기술 수준이 미치지 못한 상태이다.

패브릭 기반 인프라(Fabric-Based InfraStructure: FBI)는 데이터센터의 인프라 관리 방법이 개개의 장치마다 달라지는 현재의 방식이 아니라 자원 풀을 중심으로 워크로드에 따라 필요한 컴퓨팅 자원이 구성/재구성



〈자료〉: Gartner, 2010.

(그림 2) Fabric-Based Infrastructure 개념도

될 수 있는 관리 방식을 의미한다. FBI는 관리 방법의 변화이므로 Fabric Resource Pool Management(FRPM)의 역할이 중요해진다. FRPM을 통해 물리 자원을 구성하고 제공하며, 물리 자원은 패브릭 인터커넥트로 연결되어 있다. (그림 2)에서 FBI의 개념도를 도시하였다.

FBI는 서버의 구조(현재의 서버, 패브릭 가능 컴퓨터, 패브릭 기반 컴퓨터)를 구분하지 않고 데이터센터의 인프라를 관리하는 차원에서만 고려되는 것으로 보면 적절하다. 현재의 기술 수준으로도 FBI의 개념에 대한 구현 수준은 충분하며 차후에는 서버의 구조 변화에 따라 제어 가능한 모듈의 세밀도가 달라질 것으로 예상된다.

패브릭 컴퓨팅이라는 표현은 컴퓨팅 자원을 실시간으로 제공하는 인프라(Real-Time Infrastructure: RTI)를 지향하는 관점에서 출발하였으므로 FEC보다는 FBC와 FBI를 포괄하는 개념 모델로서 생각하는 편이 적절하다. 본고에서는 이 두 가지 개념을 중심으로 패브릭 컴퓨팅을 기술한다.

Ⅲ. 패브릭 컴퓨팅 요소 기술

1. 인프라 가상화 및 관리 기술

패브릭 기반 컴퓨터에서 인프라와 자원을 가상화하고 관리하는 방법은 클라우드 컴퓨팅에서의 자원 관리와 운영체제에서의 자원 관리 방법을 혼합적으로 사용하여 달성 가능하다. 데이터센터 전체를 관리하고 자원을 제공하는 측면에서 클라우드 컴퓨팅의 자원 관리 방법이 많은 부분 포함되어야 한다. 또, 다루는 자원의 단위가 클라우드 컴퓨팅보다 세밀하다는 측면에서 (프로세싱 유닛, 메모리, I/O를 관리하므로) 운영체제의 기능과 닿아 있다.

패브릭 컴퓨팅은 패브릭 인터커넥트가 구현되어 모든 자원을 효율적으로 연결하고 분할하는 것이 가능하다면 쉽게 구현될 수 있겠으나 현실적으로 모든 자원이 하나의 전송 매체로 연결되기는 어렵다. 패브릭 기반 컴퓨터는 하나의 OS가 실행되는 싱글시스템으로 구현되는 것이 효율적일 것으로 판단되지만, 반드시 싱글시스템으로 구현될 필요는 없다.

사용 가능한 싱글시스템 이미지 기술은 하드웨어 연결망 기반의 기술과 소프트웨어 기반의 기술이 있는데, NumaScale의 NumaConnect, ScaleMP의 vSMP(Versatile SMP)가 각각 대표적인 제품이다.

가. NumaScale의 NumaConnect[5]

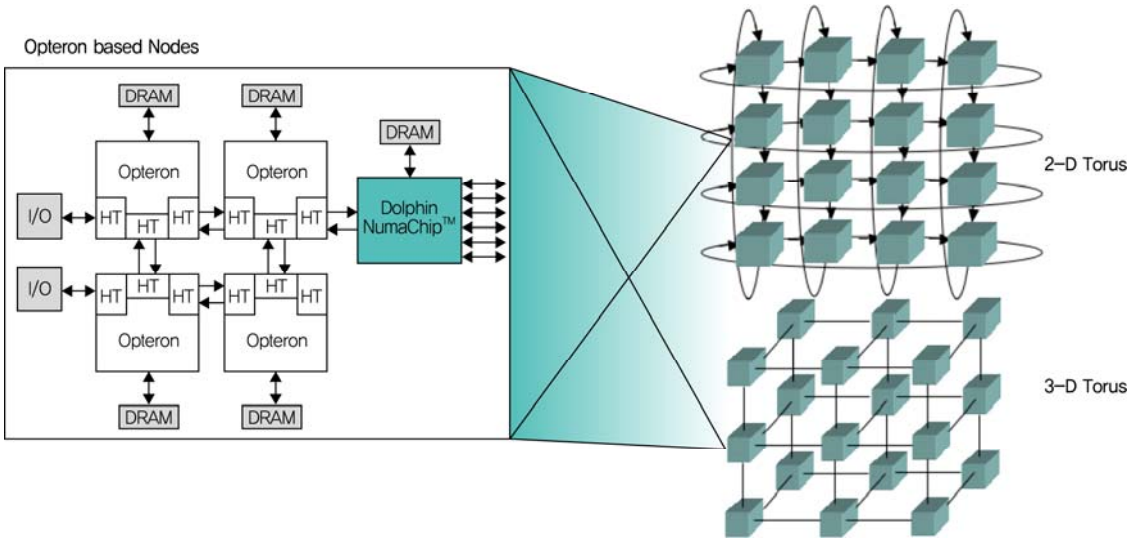
NumaConnect는 NumaConnect 어댑터와 NumaChip으로 구성되는데 낮은 비용으로 메인프레임급의 서버를 만들기 위한 기술로서 최대 4,096대의 노드까지 공유 메모리와 캐시 일관성(cache coherence)을 지원하는 싱글시스템으로 구성할 수 있다. 하드웨어에서 싱글시스템 구성에 필요한 기능을 모두 제공하므로 운영체제가 복잡해지지 않아도 되는 것이 장점이다.

노드 간 연결은 (그림 3)과 같이 2-D torus 혹은 3-D torus로 구성되고 연결 속도는 최대 19.2GB/s이다. NumaConnect 어댑터는 AMD에서 지원하는 프로세서 간 연결망인 HyperTransport에 연결된다. 지원되는 OS는 Linux, Windows Server, Solaris, Unix 등이다.

나. ScaleMP의 vSMP

여러 개의 컴퓨터 노드를 연결하여 싱글시스템 이미지를 구성하는 소프트웨어 기반의 접근 방법이다. 역시 캐시 일관성을 지원하며 공유 I/O를 활용하여 물리적으로 구분된 다른 노드의 I/O를 사용할 수 있다.

vSMP Foundation은 최대 128개의 노드를 연결할 수 있으며 최대 256TB의 메모리를, 32,768개의 코어를 연결하여 사용할 수 있다. vSMP Foundation Advanced Platform에서는 요구 기반 SMP, 여러 개의 작은 SMP 머신으로 나누는 파티셔닝, 128개 이상의 시스템 연결을 지원한다.



(그림 3) NumaConnect 연결 구조

2. 연결망 및 I/O 가상화 기술

현재의 컴퓨터와 구조적으로 유사하지만, 패브릭 컴퓨터는 새로운 종류의 연결망을 요구한다. 저지연 고대역폭, 빠른 스위칭이 가능해야 하며 확장성이 있어야 한다. 많은 종류의 연결망 기술이 개발되어 있지만, 패브릭 컴퓨팅을 위한 연결망 기술은 아직 부족하다.

본 절에서는 컴퓨팅 자원 연결에 사용되는 연결망 기술을 정리해 보고, 각 기술의 차이점, 패브릭 컴퓨팅에 적용 가능성, 각 연결망 기술의 가상화 지원 기능 등을 살펴본다.

가. QuickPath Interconnect(QPI)[6]

마이크로프로세서 간 또는 마이크로프로세서와 칩셋과의 외부 접속을 위해 인텔이 개발한 버스 프로토콜이다. 패킷 기반의 점대점(point-to-point) 상호 연결 버스로서 최대 25.6GB/s(3.2GHz의 클럭 속도일 때)의 속도로 전송이 가능하다. AMD의 HyperTransport(HT)에 대항하여 만들어졌으며 네할렘 아키텍처를 사용하는 인텔 코어 i7, 아이테니엄의 투킬라 프로세서부터 사용되었다.

컴퓨터 시스템에서 범용의 연결망으로 사용하기에는 확장성이 부족하지만, 프로세서 간 연결을 통해 프로세서의 고집적화의 용도로 사용할 수 있다.

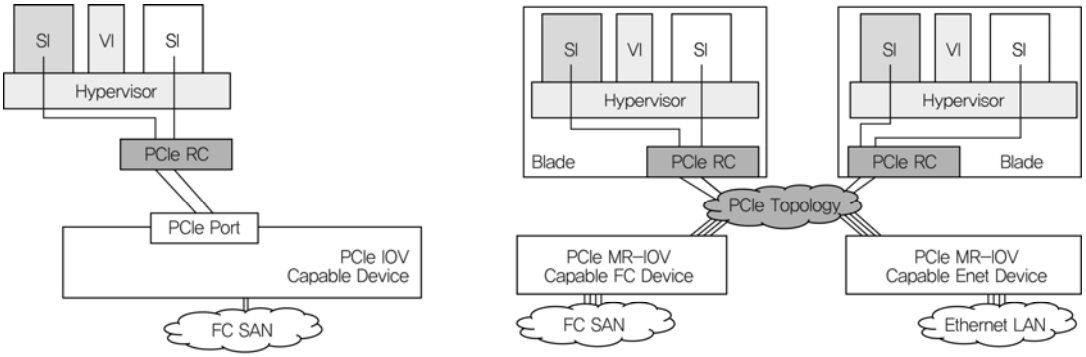
나. HyperTransport[7]

AMD, Alpha Processors, API Networks로 구성된 HyperTransport 컨소시엄에 의해 개발된 HyperTransport는 QPI와 유사하게 프로세서 간 또는 프로세서와 칩셋 간의 연결을 위해 만들어진 연결망의 일종이다. 최대 51.2GB/s(3.2GHz, 버전 3.1)의 속도로 데이터 전송이 가능하며 점대점 연결 방식이다.

범용의 연결망으로 사용할 수는 없으나, NumaConnect와 연결할 수 있으며 이 경우 많은 수의 컴퓨터 노드를 연결하여 싱글시스템을 구성할 수 있다.

다. PCI Express(PCIe)[8]

PCI-SIG(Peripheral Component Interconnect-Special Interest Group)에서 정의한 I/O 장치에 대한 연결망으로서 널리 사용되는 I/O 버스 규격이다. 현재 버전 3.0에 대한 규격이 완료되었으며 16lane의 경우 최대 32GB/s의 대역폭을 제공한다.



〈자료〉: PCI-SIG, 2008.

(그림 4) SR-IOV와 MR-IOV

PCI Express와 관련하여 살펴봐야 할 사항은 I/O 가상화이다. 현재 I/O 가상화를 제공하는 방법은 하이퍼바이저에서 네트워크 인터페이스나 스토리지 연결 어댑터에 대한 분리, 대역폭 스케줄링을 하여 소프트웨어로 가상화를 제공해 왔다. 그러나, 이런 방법이 속도가 느리고 오버헤드가 많이 들어 하드웨어에서 제공할 수 있는 가상화 방법을 고안되었는데, 그것이 (그림 4)에 표현된 SR-IOV(Single-Root Input Output Virtualization)와 MR-IOV(Multi-Root Input Output Virtualization)이다.

SR-IOV는 여러 개의 게스트 OS(혹은 System Image: SI)마다 각각 가상의 네트워크 인터페이스(vNIC(Network Interface Card))나 스토리지 어댑터(vHBA(Host Bus Adapter))를 제공하고, 동시에 접근할 수 있도록 함으로써 가상화에 따른 성능 저하를 감소시킨다.

MR-IOV는 PCIe 스위치로 연결된 여러 개의 노드(혹은 블레이드)가 여러 개의 어댑터를 동시에 접근할 수 있는 기능을 지원한다. 상황에 따라 동적으로 I/O 자원을 할당하거나 사용할 수 있게 함으로써 패브릭 기반 컴퓨터에서 지원하고자 하는 I/O 가상화 요구를 만족하는 I/O 가상화 방법이다. 제품으로 구현된 사례가 아직 없으나 실현될 경우 물리 자원을 동적으로 제어하기 위하여 다양하게 사용할 수 있을 것으로 판단된다.

라. InfiniBand[9]

PCI를 대체하기 위해 개발된 InfiniBand는 고성능컴퓨팅 시스템과 기업용 데이터센터를 위한 연결망으로 많이 사용된다. 스위치 기반의 점대점 연결망 구조를 갖고 있으며 2.5Gb/s(4wire)의 양방향 대역폭을 제공한다. 흔히 이더넷이 제공하기 어려운 빠른 속도와 낮은 지연시간을 필요로 하는 컴퓨터 환경에 사용되어 왔다.

스위치 패브릭(switched fabric)이라는 네트워크 토폴로지로 연결되는데, 하나의 노드가 하나 이상의 네트워크 스위치에 연결한다. 매우 높은 확장성을 제공하며 다른 연결망에 비해 유연한 것이 특징이다.

Ⅳ. 제품 개발 동향

패브릭 컴퓨팅을 제공한다는 기치 아래 몇 개의 제품이 소개되고 있으나, 현재까지는 패브릭 기반 인프라(FBI)에 해당하는 수준의 제품에 머무르고 있다. 관리 소프트웨어 패키지를 통해 서버, 네트워크 자원, 스토리지를 관리할 수 있으며 이를 활용하여 효율적인 데이터센터 구축과 유지를 지원하는 것이 이런 제품군의 목적이다.

패브릭 인프라 관리와 관련된 몇 가지 제품을 정리하

면 다음과 같다.

- IBM의 BladeCenter Open Fabric 제품은 블레이드 서버를 위한 관리 솔루션으로서 블레이드 서버를 일관성 있게 모니터링하고 빠르게 교체할 수 있는 환경을 지원한다.
- HP의 CloudSystem Matrix는 다양한 데이터센터 자원을 연결하여 중앙집중 관리할 수 있는 단위로 만들어 주어 클라우드 서비스 배포와 일관성과 신뢰성 유지를 위한 솔루션을 제공한다.
- Cisco의 Unified Computing System(UCS)은 하나의 블레이드 섀시에 서버와 네트워크 장비가 모두 설치되는 형태의 시스템으로 서버 설치 및 관리 비용을 줄여줄 수 있다. 또, Fabric Manager를 통해 스토리지와 네트워크를 함께 관리할 수 있는 솔루션을 개발하였으며, Nexus 1000이라는 소프트웨어 패키지를 통해 가상 이더넷과 네트워킹 정책 등을 통합 관리할 수 있는 방법을 제공하고 있다.
- Egenera에서 HP, Dell, Fujitsu 등의 블레이드 서버를 지원하는 PAN Manager라는 솔루션을 출시하여 다양한 하드웨어 자원을 하나의 플랫폼으로 관리하는 방법을 제시하였다.

V. 결론

패브릭 컴퓨팅, 특히 동적인 패브릭 기반 컴퓨터(FBC)가 구현되기 위해서는 여전히 기술적으로 해결해야 할 과제가 많다. 컴퓨터를 구성하는 모든 구성 요소가 분리되고 패브릭 인터커넥트를 통해 즉시 조합 가능한 형태로 연결되어야 한다는 지향점 때문에 더욱 그러하다.

CPU와 메모리를 연결하는 시스템 버스와 메모리에서 I/O 장치로 연결되는 네트워크의 종류는 그 특성이 다르다. 시스템 버스는 수 ns의 응답 속도와 수 GB/s의 대역폭이 요구되는데 반해, I/O 장치를 연결하는데 사용되는 네트워크는 그 정도의 속도가 필요치 않다. 현재의 기술로는 이런 속도를 달성하는 네트워크로 연결한다 하더라도 확장성(scalability)에 문제가 있어서 데이터센터 네트워크로의 확장은 불가능하다.

이런 이유로, CPU와 메모리는 직접 결합하는 방식을 취하고 메모리와 I/O 장치 연결과 그 외 연결망은 동일한 네트워크 패브릭을 통해 이루어질 것이라는 전망을 하는 경우가 많다[10].

많은 컴퓨터 전문가들이 데이터센터 전체가 단일 종류의 네트워킹 인프라로 연결되고 서버 박스에서 벗어나 분리된 개별 자원들이 실시간으로 필요에 따라 연결되는 패브릭 컴퓨팅을 미래의 컴퓨터 구조로 예견하고 기다려 왔다. 본고에서 살펴본 것처럼 아직 패브릭 컴퓨팅은 당장 사용할 수 있는 단계의 기술로 진화하지는 못하였다. 그러나, 패브릭 컴퓨팅이라는 비전을 통해 새로운 기술이 지속적으로 등장하고 있으며, 하드웨어와 시스템 소프트웨어 기술이 진화함에 따라 조금씩 가까워질 것으로 기대한다.

용어해설

싱글시스템 여러 개의 컴퓨터 시스템을 하나의 컴퓨터 시스템인 것처럼 보여주는 클러스터. 일반적으로 하나의 운영 체제가 수행되는 것처럼 사용자에게 보여줌으로써 자원의 제약을 극복하게 만들어주는 시스템

Fabric Resource Pool Management 패브릭 기반 인프라를 관리하기 위해 CPU, 메모리, I/O 장치를 자원 풀로 관리하면서 실시간으로 물리 머신을 구성, 제공하기 위한 관리 소프트웨어

약어 정리

CPU	Central Processing Unit
FRPM	Fabric Resource Pool Management
FBC	Fabric-Based Computer
FBI	Fabric-Based Infrastructure
FEC	Fabric-Enabled Computer
HBA	Host Bus Adapter
HT	HyperTransport
I/O	Input/Output

MR-IOV	Multi-Root Input Output Virtualization
NIC	Network Interface Card
OS	Operating System
PCIe	PCI Express
PCI	Peripheral Component Interconnect
PCI-SIG	Peripheral Component Interconnect-Special Interest Group
QPI	QuickPath Interconnect
RTI	Real-Time Infrastructure
SI	System Image
SR-IOV	Single-Root Input Output Virtualization
UCS	Unified Computing System
vSMP	Versatile SMP

참고문헌

[1] S. Oritiz Jr., "Stitching Computing Systems into a Single Fabric," *IEEE Comput.*, vol. 41, no. 9, Sept. 2008.

[2] 김진미 외, "차세대 컴퓨팅을 위한 가상화 기술," 전자통신동향분석, vol. 23, no. 4, 2008. 8, pp. 102-114.

[3] Wikipedia, Fabric computing. http://en.wikipedia.org/wiki/Fabric_computing

[4] J. Skorupa et al., "Clearing the Confusion about Fabric-Based Infrastructure: a Taxonomy," Gartner Research, Sept. 2010.

[5] E. Rustad, "NumaConnect: A High Level Technical Overview of the NumaConnect Technology and Products," White Paper, Numascale. http://www.numascale.com/numa_pdfs/numaconnect-white-paper.pdf

[6] Intel, "An Introduction to the Intel QuickPath Interconnect," Jan. 2009. http://www.intel.com/ content/dam/doc/white-paper/quick-path-interconnect-introduction-paper.pdf

[7] HyperTransport™ Consortium, "HyperTransport™ I/O Technology Overview: An Optimized, Low-latency Board-level Architecture," White Paper, June 2004. http://www.hypertransport.org/docs/wp/HT_Overview.pdf

[8] M. Wagh, "IOV Overview and Update," PCI-SIG, 2008.

[9] Mellanox Technologies, "Introduction to InfiniBand," 2003. http://www.mellanox.com/pdf/whitepapers/IB_Intro_WP_190.pdf

[10] A. Ingram, "Fabric Computing: Gartner's View for the Future of the Datacenter," Feb. 2011. http://forums.juniper.net/t5/The-New-Network/Fabric-Computing-Gartner-s-view-for-the-future-of-the-data center/ba-p/75820