

# 가상화와 컴퓨터 네트워크의 활용 2: 데이터센터 네트워크 구성 - (1)

최영락 휴레이포티지브 선임연구원 & 오픈플로우코리아 기술매니저

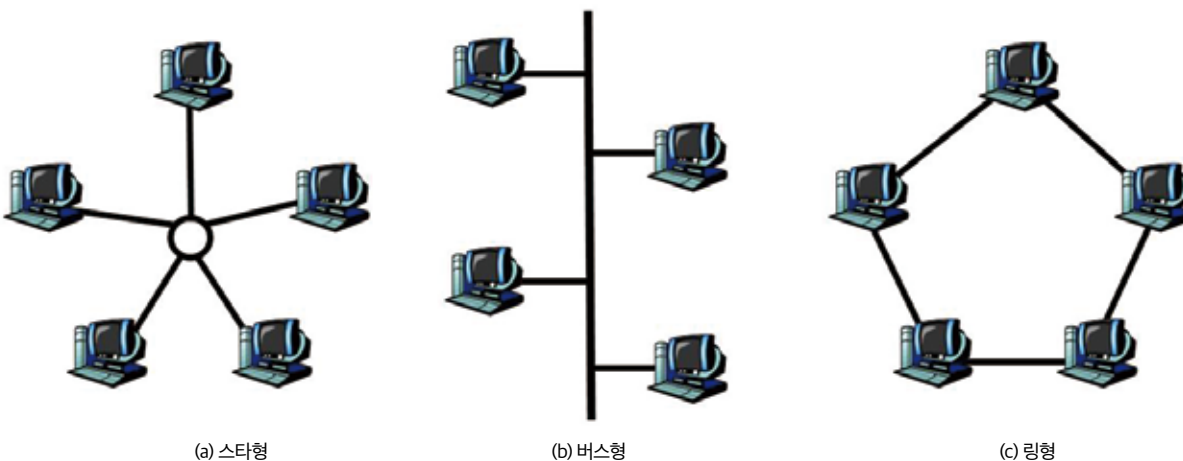
- 연재 목록 -

- |                               |                       |
|-------------------------------|-----------------------|
| 1. 가상화와 데이터센터                 | 6. 네트워크 스토리지 기술 - (2) |
| <b>2. 데이터센터 네트워크 구성 - (1)</b> | 7. 활용 사례 - (1)        |
| 3. 데이터센터 네트워크 구성 - (2)        | 8. 활용 사례 - (2)        |
| 4. 가상화와 네트워크 스토리지             | 9. 활용 사례 - (3)        |
| 5. 네트워크 스토리지 기술 - (1)         |                       |

지난 '가상화와 컴퓨터 네트워크의 활용' 첫 번째 연재에서는 가상화에 대해 살펴보고, 가상화와 관련된 하이퍼바이저, 클라우드 컴퓨팅, 그리고 데이터센터의 트래픽 변화에 대해 살펴보았습니다. 이번 연재에서는 데이터센터의 네트워크 구성에 대해 관련 용어들과 함께 설명하고, 데이터센터 네트워크의 토폴로지에 대해 살펴보려고 합니다.

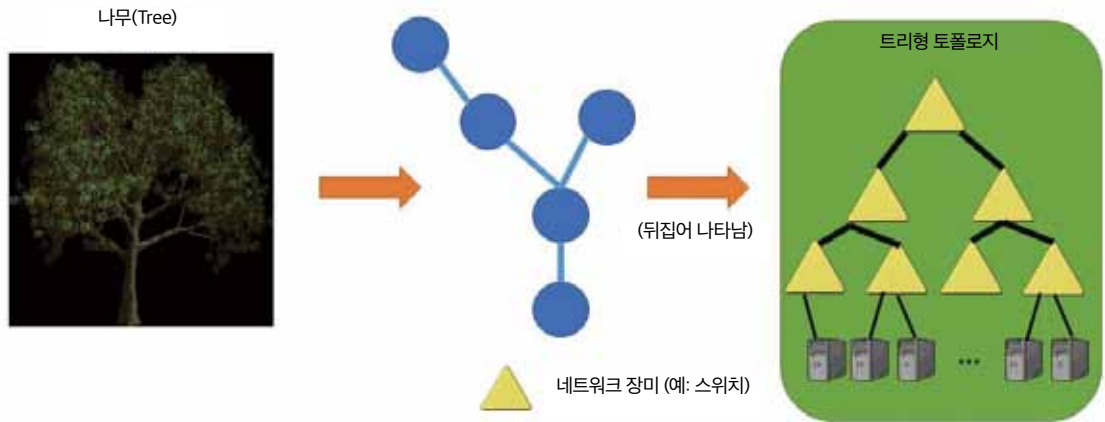
### 데이터센터 네트워크 관련 용어

일반적으로 컴퓨터와 네트워크 장비, 그리고 이들이 서로 연결되어 있는 선들을 중심으로 하여 특정 환경에 대해 어떻게 배치되어 있는지 나타낸 그림을 토폴로지(Topology)라고 합니다. 즉, 네트워크의 배열 및 구성에 대해 스위치, 라우터 등을 포함한 네트워크 장비, 그리고 네트워크를 사용하는 서버들에 대한 구성도를 이야기합니다. 이러한 토폴로지는 컴퓨터 네트워크에서 각 장비들이 서로 연결되는 방식과 관련이 있기에, 이전 '컴퓨터 네트워크의 이해 및 활용 - 데이터 물리 계층'에서 살펴보았던 네트워크 연결 방식과 밀접하게 관련이 되어 있습니다. 당시 스타형, 버스형, 링형 3가지를 살펴보았었는데, 이번 연재에서는 전통적인 데이터센터 네트워크 구조에서 주로 사용하던 트리형 토폴로지에 대해 살펴보려고 합니다.



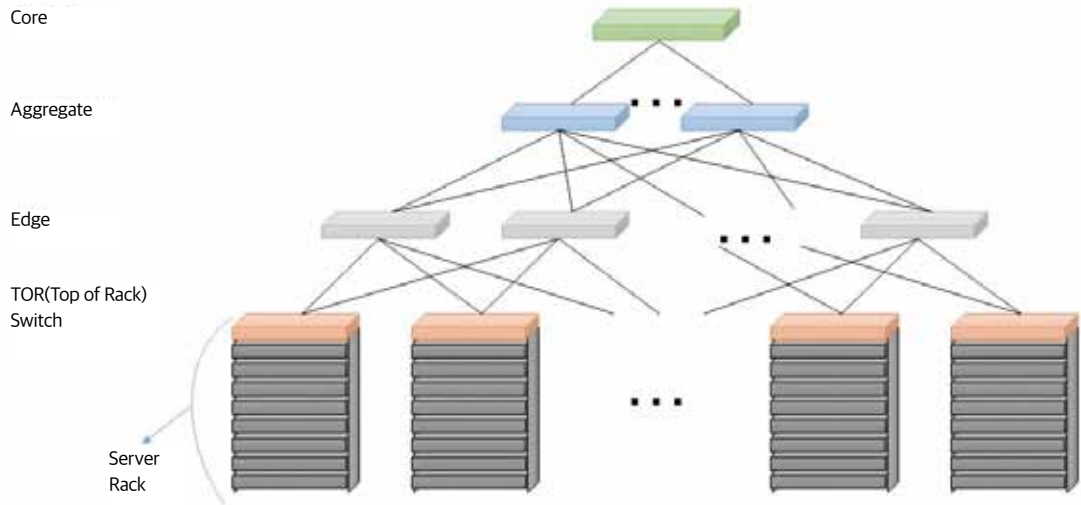
[그림 1] 데이터링크 계층 - 네트워크 연결 방식

트리(Tree)형 토폴로지에서는 '트리'를 우리말로 직역하면 나무가 됩니다. 사실 트리형이라는 용어는 [그림 2]와 같이 비유적으로 사용된 단어입니다. 일반적으로 나무는 하나의 뿌리에서 자란 후, 여러 가지가 생기고, 가지 위에 꽃이나 열매를 맺게 됩니다. 네트워크 토폴로지 형태가 이러한 나무 형상을 뒤집었을 때와 유사하다고 하여 트리형 토폴로지라는 용어를 사용합니다.



[그림 2] 트리형 토폴로지 개념

[그림 3]은 이러한 트리형 토폴로지 형태를 따르는 전통적인 데이터센터 토폴로지 개념도에 해당합니다. 해당 개념도에서 각 위치에 있는 장비들은 다음과 같습니다.

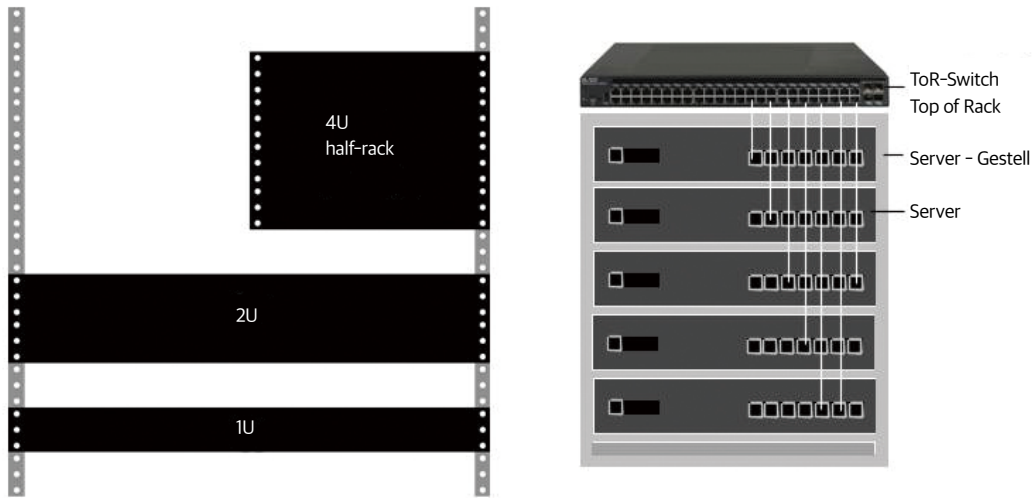


[그림 3] 전통적인 데이터센터 토폴로지 개념도

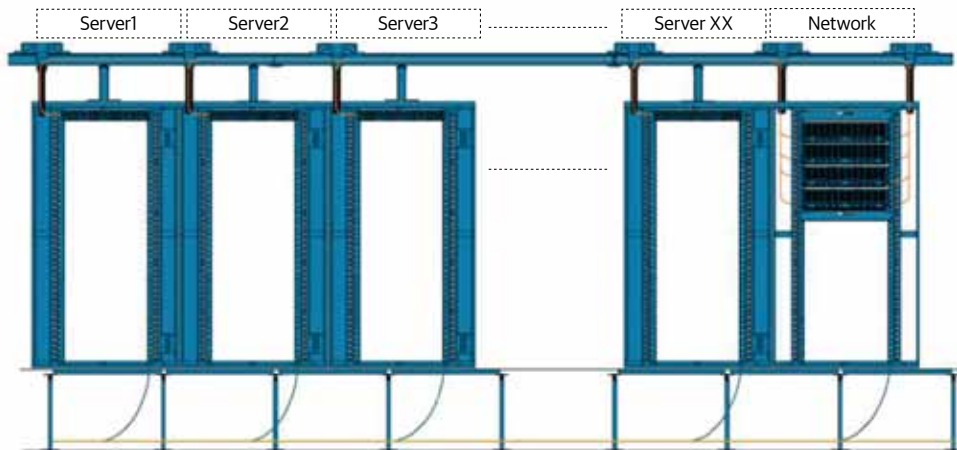
**서버 랙(Server Rack)** : 일반적으로 데이터센터에는 랙(Rack)이라는 규격화된 서버 장비들이 설치됩니다. 보통 가로 19인치 길이를 기준으로 하고, 세로는 1.75인치를 기준으로 유닛(Unit, 약어로 U를 사용) 단위로 장비가 제작됩니다. 예를 들어, 1U 랙 장비이면 가로 19인치 × 세로 1.75인치 크기로 랙에 설치 가능한 장비입니다.

**TOR(Top of Rack) 스위치** : 서버들이 설치되어 있는 서버 랙 상단에 배치되어 있는 스위치로, 해당 랙에 설치된 서버들에 대한 트래픽을 수용하기 위해 배치된 스위치입니다.

**Edge, Aggregation, Core 계층** : 여러 개의 TOR 스위치들에 대한 트래픽을 수용하기 위해서 상단에는 TOR 스위치보다 많은 양의 트래픽을 수용 가능한 고가의 네트워크 장비들을 배치합니다. 해당 장비들은 데이터센터 네트워크 상황에 따라 스위치가 될 수도 있고, 라우터가 될 수도 있습니다. 예를 들어 TOR 스위치가 1Gbps를 수용하는 스위치 장비라면, Edge나 Aggregation 계층에는 10Gbps를 수용하는 네트워크 장비가 됩니다. 이러한 네트워크 장비들은 제공 / 설치 업체에 따라 사용하는 용어들이 조금씩 다를 수 있습니다. 다른 데이터센터 토폴로지 그림들을 보면 Edge 계층 없이 Aggregation과 Core 계층만 존재하는 그림도 있으며, Edge 대신 Access 라는 용어를 사용하여 표현하기도 합니다. 또는 여러 개의 서버 랙들이 있다면 네트워크 랙 또한 별도로 위치시키는 경우 해당 네트워크 랙에 TOR 스위치들에 대한 트래픽을 허용하는 고급 네트워크 장비들이 설치되기에 해당 네트워크 장비들을 한 줄의 끝 부분에 있다고 하여 EOR(End of Row)에 있는 장비라고 이야기하기도 합니다.

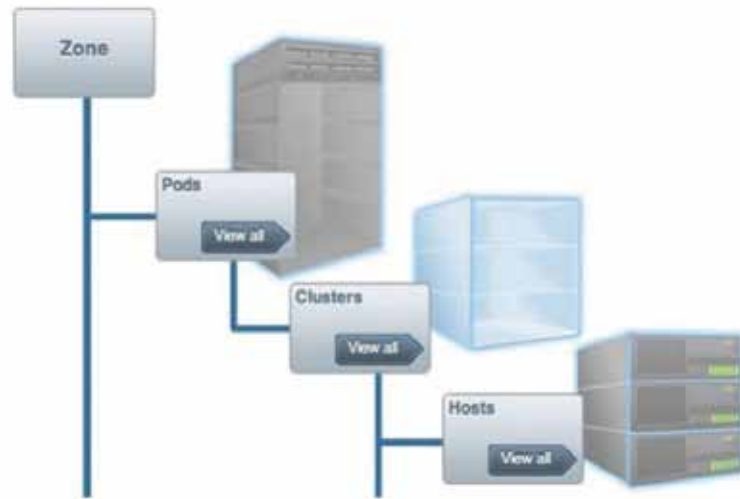


[그림 4] Rack 및 TOR 스위치 구성 / 출처 : Wikipedia, itwissen.info



[그림 5] EOR 스위치 구성 (오른쪽 Network 랙 부분)  
출처 : Cisco, "Data Center Top-of-Rack Architecture Design"

이러한 데이터센터는 클라우드와 같은 서비스 관점에서 논리적으로 구분한 용어를 사용하기도 합니다. 존과 팟이 대표적인 용어인데, 우선 팟(Pod)은 Point of Delivery의 약어로, 컴퓨팅이 가능한 서버를 포함하여 네트워크, 스토리지 및 응용프로그램까지 묶어 기본적인 네트워크 서비스를 제공 가능한 기본 단위를 이야기합니다. 이에 대한 데이터센터 물리 구조에서는 보통 1개의 랙을 의미하지만, 때로는 2~3개의 랙을 묶어 1개의 팟으로 보는 경우도 있습니다. 반면, 존(Zone)은 데이터센터에서 큰 규모의 서비스 묶음을 제공 가능한 단위를 의미합니다. 소규모 데이터센터에서는 데이터센터 전체가 1개의 존을 의미하기도 하고, 데이터센터 규모가 클수록 한 데이터센터 내에 여러 개의 존을 포함하게 됩니다. 그리고 논리적인 1개 팟 내에는 여러 호스트들이 존재하는데, 해당 호스트들을 묶어 클러스터 단위로 부르기도 합니다.

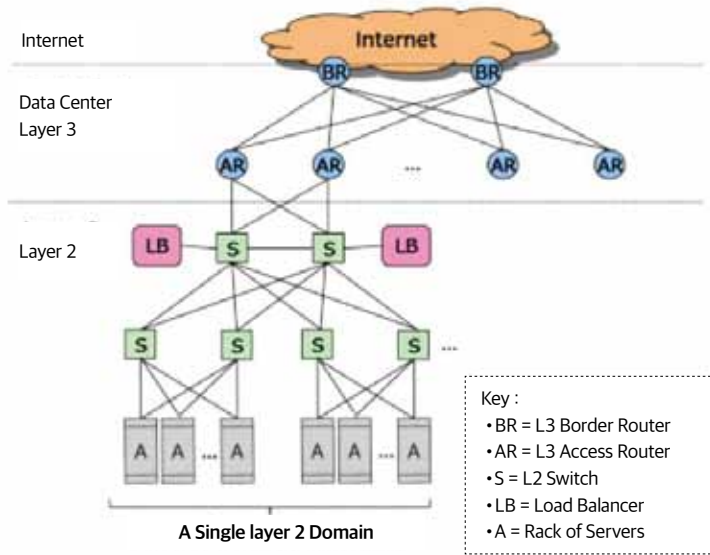


[그림 6] 데이터센터 논리적인 단위 / 출처 : CloudStack

### 데이터센터 네트워크 토폴로지의 특징

데이터센터 네트워크 토폴로지에는 여러 특징들이 있습니다. 먼저 살펴볼 특징은 오버-서브스크립션(oversubscription)이라는 특징입니다. [그림 3]과 같은 토폴로지를 가진 전통적인 데이터센터 네트워크에서는 TOR 스위치를 시작으로 상위 네트워크로 갈수록 수용하는 트래픽 양이 점차적으로 늘어날 수밖에 없습니다. 즉, 비용 및 효율성을 고려하여 1개의 상위 계층의 네트워크 장비는 여러 개의 하위 계층 네트워크 장비들을 수용하게 되는데, 이에 대한 비율을 오버-서브스크립션이라고 합니다. 예를 들어, 2:1이면 1개의 상위 계층 네트워크 장비가 평균 2개의 하위 계층 네트워크 장비를 수용하는 식입니다. 이 오버-서브스크립션 비율에 대해 여러 실무 등을 통한 일반적인 가이드라인에 따르면 2.5:1에서 8:1 값을 권장합니다만, 전통적인 데이터센터에서는 Core 계층에 80:1, 심지어는 240:1까지 도달하는 경우도 있습니다. (많이 몰릴수록 상위 계층에 부담이 크다는 것을 의미하며, 패킷 손실이나 딜레이와 같은 문제가 발생할 수 있습니다)

또한 [그림 3]에서처럼 가장 많은 트래픽이 몰리는 Core 계층에 1개의 네트워크 장비만을 사용하는 경우에는 해당 장비에 장애가 발생하는 경우 데이터센터 네트워크 전체가 마비될 것입니다. 따라서 전통적인 데이터센터 네트워크에서는 [그림 3]과 같은 토폴로지 구조를 기반으로 같은 종류의 장비들을 2대 이상 배치하여 특정 장비에 장애가 발생하였을 때 대처 가능하도록 구성합니다. [그림 7]은 Core와 Access 계층에 장비들을 이중으로 두고 서로 연결하여 각 계층에서 어느 1대의 네트워크 장비에 장애가 발생하더라도 네트워크 기능이 정상적으로 동작하는 전통적인 데이터센터 네트워크 토폴로지 예시입니다.



[그림 7] 이중화된 데이터센터 네트워크 토폴로지 예시 / 출처 : VL2, SIGCOMM 2009

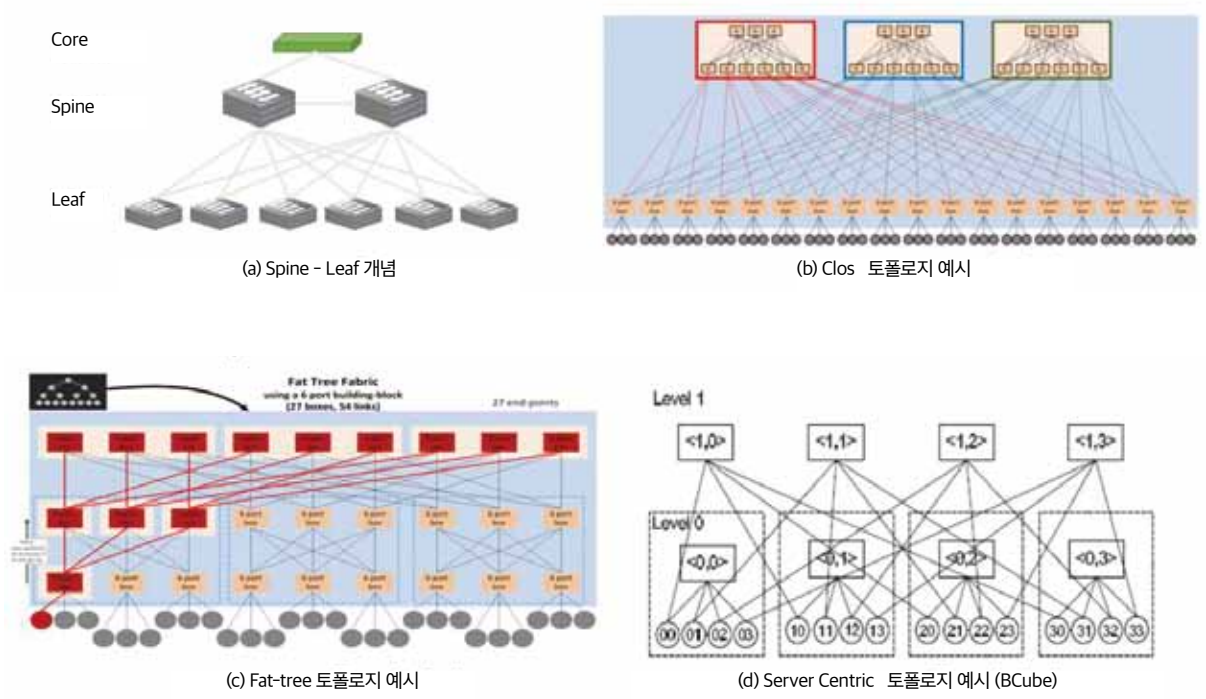
그런데 [그림 7]은 [그림 3]과 달리 데이터센터 네트워크에서 2계층(Layer 2, 데이터링크 계층, 이하 L2)과 3계층(Layer 3, 네트워크 계층, 이하 L3)을 구분 지었습니다. L2 네트워크 장비들은 MAC 주소를 살펴보는 스위치 장비들에 해당하며, L3 네트워크 장비들은 IP 주소를 고려하는 L3 스위치 또는 라우팅 기능을 수행하는 라우터 장비에 해당합니다. 이 L2와 L3에 대한 범위 크기에 따라 데이터센터 네트워크의 여러 특징에 영향을 미칩니다. 우선 L2 범위를 크게 설정하면, 하나의 L2 범위 내에 포함된 서버 랙의 수가 많아지게 되는 것입니다. 즉, 같은 L2 범위 내에 포함된 서버 수가 증가하게 되는데, 이때 브로드캐스트 범위가 커져 브로드캐스트에 따른 트래픽이 증가하는 단점이 있습니다. 하지만 같은 범위 내에서는 서버들의 IP 주소가 변경되더라도 MAC 주소 학습과 같은 방법 등을 통해 상위 L3 등의 라우팅 경로를 수정하지 않고도 그대로 사용이 가능해 집니다. 반대로, L2 범위가 작은 경우에는 L3 범위가 커져 브로드캐스트 범위가 작아지는 이점은 있으나, 라우팅 경로 관리의 복잡함, IP 주소 변경에 따른 여러 관리의 복잡성에 따른 단점이 있습니다.

### 데이터센터 네트워크 요구 사항

이러한 데이터센터 네트워크에는 점차 많은 서버들이 데이터센터에 수용되며 복잡한 응용프로그램들이 해당 서버 위에 동작하게 되었고, 무엇보다 서버 가상화에 따라 많은 가상 머신들이 물리 서버 위에서 실행되면서 많은 요구 사항들이 생겨났습니다. 다음은 해당 주요 요구 사항들을 정리한 것입니다. 모든 네트워크 요구 사항들을 만족하는 데이터센터 네트워크를 필요로 한다기보다는, 대상 데이터센터 특성을 고려하여 요구 사항의 우선순위를 정리하고, 해당 요구사항에 따른 확장성(Scalability), 경제성, 성능, 안정성, 응용 프로그램 지원 여부 등을 고려하여 데이터센터 네트워크 토폴로지 설계에 반영될 것입니다.

- ① 특정 물리 서버에 있는 임의의 가상 머신이 다른 물리 서버로 이동하더라도, IP 주소가 바뀌지 않도록 구성
- ② 가상 서버 변경에 따라 네트워크 장비들을 수동으로 설정하는 경우가 없도록 구성
- ③ 한 물리/가상 서버에서 다른 물리/가상 서버로 통신 시 가장 효율적인 네트워크 경로를 선택하여 통신하도록 구성
- ④ 해당 네트워크 토폴로지에서 트래픽 루프(Loop)가 발생하지 않도록 구성
- ⑤ 장애에 대한 빠른 탐지 및 복구가 가능할 것

지난 연재에서 데이터센터 트래픽의 패턴이 변화하였다고 하였습니다. 이러한 트래픽 패턴의 변화에 따라 해당 트래픽 패턴에 최적화된 다양한 데이터센터 토폴로지가 등장하였습니다. [그림 8-(a)]는 그중 하나의 예로, 일종의 메쉬(Mesh, 그물) 형태로 하나의 하위 계층 장비가 모든 상위 계층의 네트워크 장비에 연결하는 식으로 구성합니다. 이렇게 토폴로지를 구성하면 모든 하위 계층 장비들은 상위 계층 장비에 직접 연결되어 회선 복잡성이 증가하고 이에 따른 관리의 어려움은 있으나 East-West 트래픽을 수용하기에 보다 수월해진다는 장점이 있습니다. 이러한 네트워크 토폴로지를 Spine-Leaf 토폴로지라고 합니다. 이를 기반으로 한 Clos 토폴로지 [그림8-(b)], Fat-tree 토폴로지 [그림8-(c)] 등과 같은 다양한 데이터센터 토폴로지가 제안되고 있으며, 연구에서는 [그림 8-(d)]와 같이 각 서버에서도 네트워크 기능을 수행하도록 하는 서버 기반(Server Centric)의 데이터센터 네트워크 토폴로지가 제안되기도 하였습니다.



[그림 8] 데이터센터 네트워크 트래픽을 위한 다양한 토폴로지  
b,c 출처 : 솔라구구 블로그, <http://sola99.tistory.com/>

이와 같이 이번 연재에서는 TOR 스위치, Edge-Aggregation-Core 등의 데이터센터의 네트워크 관련 용어들과 함께 데이터센터 네트워크 토폴로지의 특징을 L2와 L3 등으로 구분 지어 살펴보았습니다. 다음 연재에서는 데이터센터 네트워크에 대해 네트워크 가상화라는 개념과 같이 계속 살펴볼 예정입니다. 📖