

지난 10년 동안 혁신적인 클라우드 제공업체들이 어플리케이션 제공 비용을 크게 낮춘 이후로, 업무 속도와 비용 효율성을 높이기 위해 전 세계적으로 클라우드 도입률이 높아지고 있습니다. Arista의 획기적인 클라우드 네트워킹 솔루션은 이러한 클라우드화를 가능하게 만들었습니다. 소셜 미디어/빅데이터 등의 새로운 최신 어플리케이션, 고밀도 서버 가상화, IP 스토리지 등의 새로운 아키텍처, 그리고 모든 어플리케이션에 모바일로 액세스할 수 있어야 한다는 요구 사항으로 인해 데이터센터의 네트워크 인프라에 대한 수요가 매우 높아졌습니다. 클라우드 도입을 가능하게 만드는 네트워크 운영 체제와 네트워크 아키텍처는 오버-서브스크립션이 적용되고 비용이 많이 드는 과거의 다중 계층의 계층형 레거시 솔루션과는 근본적으로 다릅니다. 더 높은 대역폭이 필요한 고성능 서버와 어플리케이션의 도입률이 증가함에 따라, 10/25 Gigabit Ethernet 스위칭과 40/100 Gigabit Ethernet을 함께 도입하는 추세입니다. 최신 세대의 스위치 실리콘을 사용하면 10/40 Gigabit Ethernet에서 25/100 Gigabit Ethernet으로 원활하게 전환할 수 있습니다. 이 백서에서는 Arista의 2 계층 스파인/리프 및 단일 계층 Spline™ 범용 클라우드 네트워크 설계에 대해 자세히 설명합니다. 이러한 설계는 벤더 전용 프로토콜, 벤더 종속성 또는 복잡한 물리적 업그레이드 없이 최고의 확장성, 성능, 집적도를 제공합니다.

ARISTA 설계의 핵심 요소

모든 Arista 범용 클라우드 네트워크 설계에 적용되는 9 가지 핵심 설계 목표는 다음과 같습니다.

1. **전매 프로토콜 또는 벤더 고착화 없음.**Arista는 개방형 표준을 지지하며, 우수성이 입증된 Arista의 업계 입증된 레퍼런스 디자인들은 벤더 전용의 프로토콜이나 벤더 종속성 없이도, 도메인 내 대규모 결함이 없고 집중성이 높은 초대형 확장 네트워크를 구축할 수 있음을 보여 줍니다.
2. **적을수록 좋은 계층 수.**몇 년 전까지만 해도 레거시 네트워크에서 필요한 포트 수를 확보하려면 3 개 이상의 계층이 필요했지만, 이제는 1 계층 또는 2 계층 설계에서도 동일한 수의 포트를 사용할 수 있습니다. 계층 수가 적은 설계(예: 3 계층이 아닌 2 계층 스파인/리프 설계)를 적용하는 경우 비용, 케이블, 전력/발열을 줄이고 네트워크를 단순하게 유지할 수 있습니다. 단일 계층 Spline 네트워크 설계에서는 스위치 계층을 상호 연결하는 데 포트를 사용하지

않으므로 사용 가능한 포트당 비용을 최소한으로 유지합니다.

3. **다양한 프로토콜 사용 가능.**Arista는 L2의 유연성과 L3의 확장성 및 복구 특성을 모두 가진 VXLAN 등의 개방형 멀티 벤더 지원 프로토콜을 사용해 레이어 2/3 또는 하이브리드 L2/L3 설계에서 구축된 확장형 설계를 지원하므로, 대규모 결함 도메인에서 나타나는 근본적인 문제 없이 대형 설계를 활용할 수 있습니다.
4. **active/active 상태에서 최신 인프라 실행.**레이어 2의 MLAG(Multi-Chassis Link Aggregation) 및 레이어 3의 ECMP(Equal Cost Multi-Pathing)을 통해 인프라를 포트 차단 없이 active/active 상태로 구축할 수 있으므로 네트워크는 두 장치 간에 사용되는 모든 링크를 사용할 수 있습니다.

5. **속도가 빠르고 유동적인 포트 속도 조정이 가능한 설계.**오늘날의 데이터센터에 포함된 대부분의 서버/컴퓨팅 및 스토리지는 40G 또는 100G 에서 작동하는 스위치에서 10G 업링크 속도로 연결됩니다. 이 연결 속도를 25G 로 높이는 경우 기가비트당 비용을 낮출 수 있으며, 최신 25G/50G NIC 에서 제공되는 PCIe 3 세대 속도 및 하드웨어 오프로드 기능에 맞는 속도로 연결이 가능해집니다. Arista 스위치 및 범용 클라우드 네트워크 설계에서는 리프의 경우 1G 에서 10G, 25G 로 유동적인 마이그레이션이 가능하며, 스파인-리프 연결의 경우 10G 에서 40G, 100G 로 원활하게 전환할 수 있습니다(25G/50G 브레이크아웃 옵션 제공).
6. **확장형 설계를 통해 소형 인프라에서 점진적으로 인프라 확장.**2 방향 ECMP 설계는 2-way/4-way/8-way/16-way 그리고 최대 64-way 까지 디자인으로 확장할 수 있습니다. 즉, 초기 자본을 많이 투자하지 않아도 장기적으로 ECMP 설계를 확장할 수 있습니다.
7. **중요한 대형 버퍼 활용.**최신 운영 체제, NIC(네트워크 인터페이스 카드), 확장형 스토리지 어레이는 TSO(TCP 세그먼테이션 오프로드), GSO, LSO 등의 기술을 사용합니다. 이러한 기술은 서버에서 대량의 데이터를 보낼 때 필요한 CPU 주기를 줄이는 데 기본적으로 사용됩니다. 이와 같은 기술의 단점은 일정량의 데이터를 전송하려는 어플리케이션/OS/스토리지의 경우 해당 데이터를 NIC 로 오프로드하며, 그러면 NIC 는 데이터를 세그먼트로 분할한 다음 라인 속도에서 연속 프레임으로 와이어에 보낸다는 점입니다. 이 중에서 둘 이상의 세그먼트가 같은 출력 포트에 전송되는 경우에는 마이크로버스트 혼잡이 발생합니다. 이러한 버스트를 처리하는 방법 중 하나는 버스트를 흡수할 수 있도록 링크를 최소의 오버-서브스크립션 또는 오버-프로비저닝을 적용하여 네트워크를 구축하는 것입니다. 또한 트래픽 팬인(fan-in)을 줄일 수도 있습니다. 아니면 패킷 드롭에서 발생한 버스트를 흡수하는 덩 버퍼와 함께 스위치를 배치하는 방법도 있습니다. 이 경우 유용한 처리량도 감소하게 됩니다.
8. **일관된 기능 및 OS 사용 가능.** 모든 Arista 스위치는 동일한 Arista EOS 를 사용합니다. 따라서 플랫폼, 소프트웨어 제품군 또는 OS 에 차이가 없으며 모든 스위치에서 동일한 바이너리 이미지가 사용됩니다.
9. **상호 운용성.** Arista 스위치 및 설계는 고객 고착화 없이도 다른 네트워킹 벤더의 제품과 상호 운용이 가능합니다.

설계 선택 - 계층 수

네트워크 설계에서 허용되는 원칙에 따르자면 설계 시 단기적 요구 사항이 아니라 네트워크 또는 네트워크 POD 가 시간이 경과함에 따라 어느 정도까지 규모가 커질 수 있는지에 대한 장기적인 요구 사항을 기준으로 삼아야 한다는 것입니다. 즉, 사용 가능한 최대 필수 포트 수와 장기적으로 해당 포트에 연결할 장치 간의 트래픽에 대해 원하는 초과 사용 비율을 기준으로 네트워크 설계를 선택해야 합니다.

포트 수에 대한 장기적 요구 사항을 단일 스위치나 HA 설계의 스위치 쌍으로 충족 가능하다면 단일 계층 Spline 설계를 사용하면 됩니다.

SPLINE 네트워크 설계

Spline 설계에서는 기존의 *스파인* 계층과 *리프* 계층이 하나의 Spline 으로 축소됩니다. 단일 계층 Spline 설계에서는 스위치 계층을 상호 연결하는 데 사용되는 포트가 없으므로 항상 자본적 지출과 운영 비용을 최소화할 수 있고 레이턴시도 가장 낮게 유지될 뿐만 아니라 관리

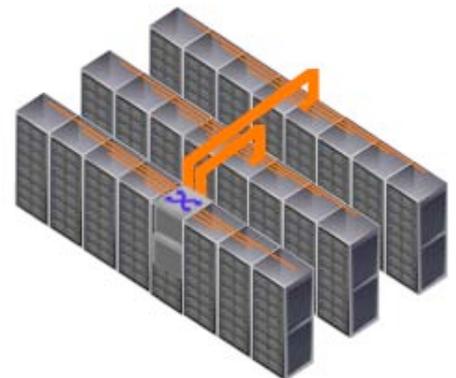
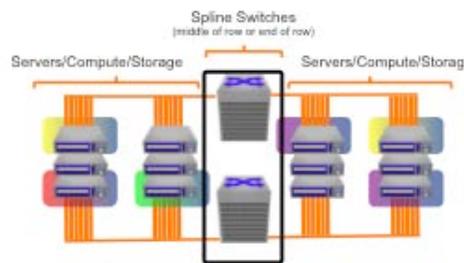


그림 1: 최대 2,000 대의 물리적 서버(1U 서버 랙 49 개)로 확장 가능한 Arista Spline 단일 계층 네트워크 설계

터치 포인트가 최대 2 개에 불과하므로 기본적으로 초과 사용되지 않습니다.

표 1 에는 단일 계층 Spline 네트워크 설계에 사용할 수 있는 Arista 스위치가 나와 있습니다. 이러한 설계의 최소 규모는 25G 또는 10G 포트 128 개(랙 3 개), 더 크게는 25G 포트 1,024 개에서 최대 10G 포트 2,048 개(랙 49 개)까지 됩니다.

표 1: Spline 단일 계층 네트워크 설계 *

스위치 플랫폼	최대 포트 수					스위치 인터페이스 유형	주요 스위치 플랫폼 특성
	10G	25G	40G	50G	100G		
Arista 7500E Series Arista 7508E Arista 7504E	1152 576		288 144		96 48	RJ45(100/1000/10G-T) SFP+/SFP(10G/1G) QSFP+(40G/4x10G) MXP(100G/3x40G/12x10G) CFP2(100G) QSFP100(100G)	2 계층 스파인/리프 설계에 가장 적합하나 Spline 설계에도 사용 가능 MXP 포트 사용 시 인터페이스 속도를 가장 유연하게 조절 가능 딥 버퍼
Arista 7320X Series Arista 7328X (Arista 7324X 포함)	1024 512	1024 512	256 128	512 256	256 128	QSFP100(100G/4x25G/2x50G) QSFP+(40G/4x10G)	대형 Spline 끝/중간 칸 설계에 가장 적합하나 2 계층 설계에서 최대 용량 10G/40G 로 스파인으로도 사용 가능
Arista 7300X Series Arista 7316X Arista 7308X Arista 7304X	2048 1024 512		512 256 128			RJ45(100/1000/10G-T) SFP+/SFP(10G/1G) QSFP+(40G/4x10G)	대형 Spline 끝/중간 칸 설계에 가장 적합하나 2 계층 설계에서 최대 용량 10G/40G 로 스파인으로도 사용 가능 RJ45 10GBASE-T 를 사용하는 경우 100M/1G/10G 간의 원활한 전환 가능
Arista 7260X 및 7060X Series Arista 7260CX-64	258	256	64	128	64	QSFP+(40G/4x10G) QSFP100(100G/2x50G/4x25G)	광학/DAC 연결을 사용하는 중형 Spline 끝/중간 칸 설계에 가장 적합
Arista 7260QX-64	2		64			QSFP+(40G)	Spline 설계용 아님
Arista 7060CX-32S	130	128	32	64	32	QSFP+(40G/4x10G) QSFP100(100G/2x50G/4x25G)	광학/DAC 연결을 사용하는 소형 Spline 끝/중간 칸 설계에 가장 적합
Arista 7250X 및 7050X Series Arista 7250QX-64	256		64			QSFP+(40G/10G 4 개)	광학/DAC 연결을 사용하는 중형 Spline 끝/중간 칸 설계에 가장 적합
Arista 7050QX-32S	96		32				
Arista 7050SX-128	96		8				
Arista 7050TX-128	96		8			RJ45(100/1000/10G-T) [TX]	
Arista 7050SX-72Q	48		6			SFP+/SFP(10G/1G) [SX]	
Arista 7050TX-72Q	48		6			QSFP+(40G/4x10G) [QX]	
Arista 7050SX-96	48		6			QSFP+(40G) [SX/TX-96]	광학/DAC 연결을 사용하는 소형 Spline 끝/중간 칸 설계에 가장 적합
Arista 7050TX-96	48		6			MXP(3x40G) [SX/TX-72]	
Arista 7050SX-64	48		4			QSFP+(40G/4x10G) [SX/TX-64]	
Arista 7050TX-64	48		4			QSFP+(40G/4x10G) [TX-48]	
Arista 7050TX-48	32		4				
Arista 7150S Series Arista 7150S-64 Arista 7150S-52 Arista 7150S-24	48 52 24		4(16) (13) (6)			SFP+/SFP(10G/1G) QSFP+(40G/4x10G)	광학/DAC 연결을 사용하는 소형 Spline 끝/중간 칸 설계에 가장 적합

* 2015 년 9 월

스파인/리프 네트워크 설계

단일 계층 Spline 설계에 적합하지 않은 설계의 경우 2 계층 *스파인-리프* 설계를 차선택으로 선택할 수 있습니다. 2 계층 설계의 경우 최상위 계층에 스파인 스위치가, 최하위 계층에는 리프 스위치가 있으며 서버/컴퓨팅/스토리지가 항상 모든 랙 상단에(고밀도 리프 스위치의 경우 매 N 개 랙 위에) 있는 리프 스위치와 2 개 이상의 스파인 업링크 스위치에 연결되어 있습니다.

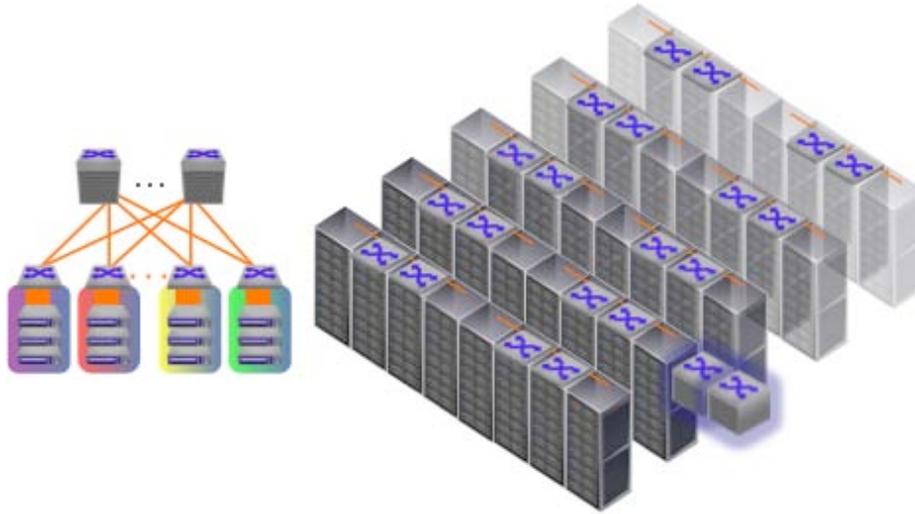


그림 2: 10 만 개가 넘는 물리적 서버를 포함하도록 확장할 수 있는 Arista 스파인/리프 2 계층 네트워크 설계

확장형 설계를 시작할 때는 스파인 스위치 1 쌍과 리프 스위치 몇 개를 포함할 수 있습니다. 10G 로 연결된 장치에 대한 초과 사용 비율이 3:1 인 2 계층 리프/스파인 네트워크 설계에는 서버/컴퓨팅/스토리지용 10G 포트 96 개와 리프 스위치당 40G 업링크 8 개가 포함됩니다. 예를 들어 Arista 7050SX-128 의 경우 10G 포트가 96 개이고 40G 업링크가 8 개이므로 초과 사용 비율이 3:1 입니다.

10G 의 경우 100G 업링크를 사용한 설계도 사용 가능합니다. 예를 들어 Arista 7060CX-32S 의 경우 서버/컴퓨팅/스토리지용 10G 포트 96 개와 100G 업링크 8 개에 대해 10G 브레이크아웃 4 개를 사용하는 100G 포트 24 개를 통해 실행됩니다. 이러한 설계의 경우 초과 사용 비율은 1.2:1 에 불과합니다.

25G 연결 장치용 설계에서는 7060CX-32S 를 사용할 수 있습니다. 이 경우 100G 포트 24 개가 서버/컴퓨팅/스토리지용 25G 포트 96 개로 브레이크아웃되며, 나머지 업링크용 100G 포트 8 개의 초과 사용 비율도 역시 3:1 입니다($25G \times 96 = 2,400G$: $100G \times 8 = 800G$).

2 계층 스파인/리프 네트워크 설계에서는 시간이 경과하면서 리프 스위치 수가 증가함에 따라 스파인 스위치 수도 동일하게 증가하도록 환경을 수평으로 확장할 수 있습니다. 최대 확장 가능 비율은 스파인 스위치 밀도, 확장 가능한 규모, 원하는 초과 사용 비율의 함수입니다. 여기서 확장 가능한 규모는 각 리프 스위치의 물리적 업링크 수와 케이블의 함수입니다.

2 계층 스파인/리프 설계에서는 스파인 스위치용으로 모듈형 또는 고정 구성 스위치를 사용할 수 있습니다. 단, 스파인 스위치를 선택하면 설계의 최대 확장 가능 범위가 고착됩니다. 표 2(리프에 대한 10G 연결)와 표 3(리프에 대한 25G 연결)에 최대 확장 가능 범위가 나와 있습니다.

표 2: 40G 업링크를 사용하는 10G 연결 장치에 대한 Arista 2 계층 스파인/리프 설계의 최대 확장 가능 범위 *

스�파인 스위치 플랫폼	스�파인 스위치의 수(확장형)	스�파인:리프 초과 사용 비율	리프-스�파인 연결	리프 스위치 플랫폼	설계가 10G 속도에서 지원하는 최대 리프 포트 수
Arista 7504E	2	3:1	40G 8 개	Arista 7050QX-32 또는 Arista 7050SX-128	리프 36 개 x 10G 96 개 = 10G 3,456 개
	4				리프 72 개 x 10G 96 개 = 10G 6,912 개
	8				리프 144 개 x 10G 96 개 = 10G 13,824 개
Arista 7508E	2	3:1	40G 8 개	Arista 7050QX-32 또는 Arista 7050SX-128	리프 72 개 x 10G 96 개 = 10G 6,912 개
	4				리프 144 개 x 10G 96 개 = 10G 13,824 개
	8				리프 288 개 x 10G 96 개 = 10G 27,648 개
Arista 7508E	2	3:1	40G 16 개	Arista 7250	리프 36 개 x 10G 192 개 = 10G 6,912 개
	4				리프 72 개 x 10G 192 개 = 10G 13,824 개
	8				리프 144 개 x 10G 192 개 = 10G 27,648 개
	16				리프 288 개 x 10G 192 개 = 10G 55,296 개
Arista 7508E	2	3:1	40G 32 개	Arista 7304X(7300X-32Q LC 포함)	리프 18 개 x 10G 384 개 = 10G 6,912 개
	4				리프 36 개 x 10G 384 개 = 10G 13,824 개
	8				리프 72 개 x 10G 384 개 = 10G 27,648 개
	16				리프 144 개 x 10G 384 개 = 10G 55,296 개
	32				리프 288 개 x 10G 384 개 = 10G 110,592 개
Arista 7316X	2	3:1	40G 64 개	Arista 7308X(7300X-32Q LC 포함)	리프 16 개 x 10G 768 개 = 10G 12,288 개
	4				리프 32 개 x 10G 768 개 = 10G 24,576 개
	8				리프 64 개 x 10G 768 개 = 10G 49,152 개
	16				리프 128 개 x 10G 768 개 = 10G 98,304 개
	32				리프 256 개 x 10G 768 개 = 10G 196,608 개
	64				리프 512 개 x 10G 768 개 = 10G 393,216 개

표 3: 100G 업링크를 사용하는 25G 연결 장치에 대한 Arista 2 계층 스파인/리프 설계의 최대 확장 가능 범위 *

스�파인 스위치 플랫폼	스�파인 스위치의 수(확장형)	스�파인:리프 초과 사용 비율	리프-스�파인 연결	리프 스위치 플랫폼	설계가 10G 속도에서 지원하는 최대 리프 포트 수
Arista 7508E	2	3:1	100G 8 개	Arista 7060CX-32	리프 24 개 x 25G 96 개 = 25G 2,304 개
	4				리프 48 개 x 25G 96 개 = 25G 4,608 개
	8				리프 96 개 x 25G 96 개 = 25G 9,216 개
Arista 7508E	4	3:1	100G 16 개	Arista 7260CX-64	리프 24 개 x 25G 192 개 = 25G 4,608 개
	8				리프 48 개 x 25G 192 개 = 25G 9,216 개
	16				리프 96 개 x 25G 192 개 = 25G 18,432 개
Arista 7328X	4	3:1	100G 8 개	Arista 7060CX-32	리프 128 개 x 25G 96 개 = 25G 12,288 개
	8				리프 256 개 x 25G 96 개 = 25G 24,576 개
Arista 7328X	4	3:1	100G 16 개	Arista 7260CX-64	리프 64 개 x 25G 192 개 = 25G 12,288 개
	8				리프 128 개 x 25G 192 개 = 25G 24,576 개
	16				리프 256 개 x 25G 192 개 = 25G 49,152 개
Arista 7328X	16	3:1	100G 64 개	Arista 7328X(7320CX-32 LC 포함)	리프 64 개 x 25G 768 개 = 25G 49,152 개
	32				리프 128 개 x 25G 768 개 = 25G 98,304 개
	64				리프 256 개 x 25G 768 개 = 25G 196,608 개

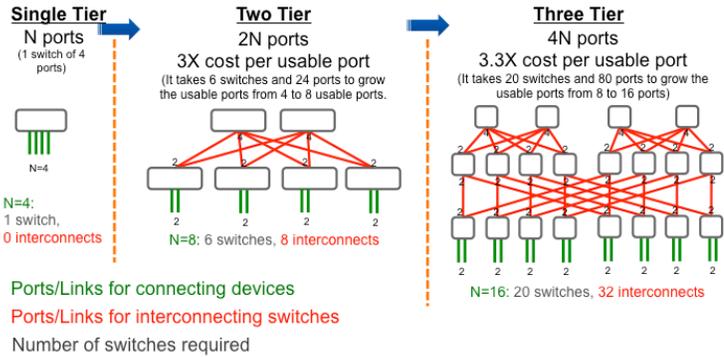
* 2015 년 9 월

리프/스파인 네트워크 설계와 관련된 설계 고려 사항

사용 가능한 포트당 자본적 지출

설계에 계층이 더 많으면 더 적은 계층에 비해 확장성이 높습니다. 그러나 자본적 지출(capex)과 운영 비용(opex)도 더 많이 듭니다. 즉, 계층이 많으면 장치 수도 많아지므로 스위치 간의 팬아웃(fan-out) 상호 연결을 위해 스위치 간에 사용되는 포트 수와 관리할 장치 수도 증가합니다.

여기서는 간단한 설명을 위해 4 포트 스위치와 초과 사용되지 않는 Clos 네트워크 토폴로지를 예로 들겠습니다. 이 네트워크에서는 포트 4 개가 필요한 경우 단일 스위치를 사용해도 요구 사항을 충족할 수 있습니다. 이 예는 간단하지만 네트워크에 적용되는 원칙을 잘 보여 줍니다.



사용 가능한 포트에 대한 요구 사항이 포트 4 개에서 8 개로 증가하며 빌딩 블록이 4 포트 스위치인 경우에는 네트워크가 단일 계층에서 2 계층으로 확장되며, 초과 사용되지 않는 네트워크를 유지하는 데 필요한 스위치 수는 1 개에서 6 개로 늘어납니다. 사용 가능한 포트 수가 2 배로 증가하면 사용 가능한 포트당 비용은 3 배로 증가합니다. 실제로 상호 연결 케이블 또는 트랜시버/파이버 비용까지 감안하면 비용이 3 배를 넘게 됩니다.

포트 수 요구 사항이 8 개에서 16 개로 다시 2 배로 증가하면 세 번째 계층이 필요하며, 스위치 수는 6 개에서 20 개로 증가합니다. 즉, 용량이 단지 2 배 증가했을 때 장치 수/비용은 3.3 배 더 증가합니다. 이 3 계층 설계는 단일 계층 설계와 비교했을 때 사용 가능한 포트가 4 배 더 많습니다(16 개와 4 개 비교). 하지만 Arista 고유의 단일 스위치 설계에 비해 비용이 20 배가 넘을 때도 마찬가지입니다.

네트워크의 규모가 확장되면 자본적 지출(capex)도 증가합니다. 그러나 더 적은 계층을 사용하여 네트워크를 구축할 수 있다면 계층 간의 상호 연결 비용이 감소하므로 자본적 지출을 크게 줄일 수 있습니다. 관리할 장치 수, 전력/냉각 정도 등이 감소하면 운영 비용(opex) 역시 크게 감소합니다. 모든 네트워크 설계는 네트워크 수명 주기 동안에 사용 가능한 포트(서버/스토리지용으로 사용되는 포트)당 비용 측면을 고려해야 합니다. 사용 가능한 포트당 비용은 다음과 같이 계산됩니다.

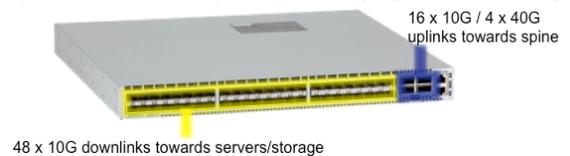
$$\frac{\text{스위치 비용(자본적 지출)} + \text{광 연결 비용(자본적 지출)} + \text{파이버 비용(자본적 지출)} + \text{전력 비용(운영 비용)}}{\text{총 노드 수} \times \text{초과 사용}}$$

초과 사용

초과 사용이란 모든 장치가 동시에 트래픽을 보내는 경우의 접속 경쟁률입니다. 이 비율은 북/남 방향(데이터센터로 유입되거나 데이터센터에서 유출되는 트래픽)과 동/서 방향(데이터센터의 장치 간 트래픽) 모두 측정 가능합니다. 대부분의 레거시 데이터센터 설계에서는 초과 사용 비율이 매우 높습니다(북/남 및 동/서 둘 모두 20:1 이상). 이는 스위치의 밀도/포트 수가 제한되어 있고 계층 수가 많으며 일반적으로 서버당 트래픽 레벨이 훨씬 낮아 왔기 때문입니다. 또한 레거시 설계의 코어 또는 집적에는 대개 L3 게이트웨이가 포함되어 있어, VLAN 간의 트래픽이 강제적으로 모든 계층을 통과해야 합니다. 여러 레벨에서 이러한 방식은 적절하지 않습니다.

멀티 코어 CPU, 서버 가상화, 플래시 스토리지, 빅데이터 및 클라우드 컴퓨팅 사용이 크게 증가하면서 최신 네트워크에 대한 요구 사항으로 낮은 초과 사용 비율이 요구되고 있습니다. 현재 사용되는 최신 네트워크 설계의 초과 사용 비율은 3:1 이하입니다. 2 계층 설계에서는 (서버/스토리지로) 다운링크 포트 대 (스파인 스위치로) 업링크 포트의 비율이 초과 사용 비율입니다. 64 포트 리프 스위치의 경우 이 비율은 48 개 다운링크 포트:16 개 업링크 포트입니다. 반면 64 포트 리프 스위치가 포함된 1:1 설계의 경우 이 비율은 32 개 다운링크 포트:32 개 업링크 포트입니다.

Typical Arista switch deployed at Top of Rack (ToR) in a Leaf/Spine design



최신 데이터센터는 초과 사용 비율을 3:1 로 시작하는 것이 좋습니다. Arista LANZ(레이턴시 분석기) 등의 기능을 사용하면 서비스 품질이 저하되기 전에(패킷 드롭으로 나타남) 혼잡이 심한 핫스팟을 파악할 수 있으므로, 트래픽이 사용 가능한 용량을 초과하는 경우 설계 비율을 어느 정도 유동적으로 수정할 수 있습니다.

그림 4:3:1의 초과 사용 비율(48x10G 다운링크 포트:4x40G 업링크 포트)로 배치된 리프 스위치

리프에서 스파인으로의 10G/40G/100G 업링크

스파인/리프 네트워크의 경우 리프에서 스파인으로의 업링크는 보통 10G 또는 40G 이며 시간이 경과함에 따라 시작점인 10G(10G N 개)에서 40G(40G N 개)로 마이그레이션할 수 있습니다. 7050SX-128 및 7050TX-128 을 제외한 모든 Arista 10G ToR 스위치에서는 속도를 이와 같이 유동적으로 변경할 수 있습니다. 이는 QSFP+가 포함된 40G 포트는 40G 1 개 또는 10G 4 개로 작동할 수 있기 때문입니다(소프트웨어 구성 가능). 또한 일부 Arista 스위치에서 제공되는 AgilePorts 기능을 사용하면 4 개의 10G SFP+ 포트에 이루어진 그룹이 하나의 40G 포트에 작동할 수 있습니다.

라인 속도에서 버스트를 발생시키는 단일 호스트의 마이크로버스트로 인해 차단되는 일이 없도록 업링크는 항상 다운링크보다 빠른 속도로 작동하는 것이 가장 좋습니다.

시간의 경과에 따른 2 계층 스파인/리프 확장

확장형 설계를 시작할 때는 대개 스파인 스위치 2 개와 리프 스위치 몇 개로 시작하게 됩니다. 이러한 설계가 시간의 경과에 따라 확장되는 방법의 예를 살펴보겠습니다. 다음 설계에는 Arista 7504E 모듈형 스위치가 스파인에, Arista 7050SX-64 가 리프에 포함되어 있으며 설계의 초과 사용 비율은 3:1 입니다. 각 리프 스위치는 서버/컴퓨팅/스토리지 연결용 10G 포트 48 개를 제공하며 스파인에 대한 총 10G 업링크 16 개를 포함하는데, 이러한 업링크는 2 개 스파인 스위치에서 10G active/active 업링크가 8 개씩 포함된 2 개 그룹으로 분할됩니다.

각 스파인 스위치에는 단일 DCS-7500E-36Q 라인 카드(40G 36 개/10G 144 개)가 포함되어 있습니다. 초기 네트워크는 그림 5 에 나와 있는 것처럼 18 개 리프 스위치에 연결이 가능하도록 확장됩니다(엔드 투 엔드 초과 사용 비율 3:1 에서 10G 연결 장치 864 개).

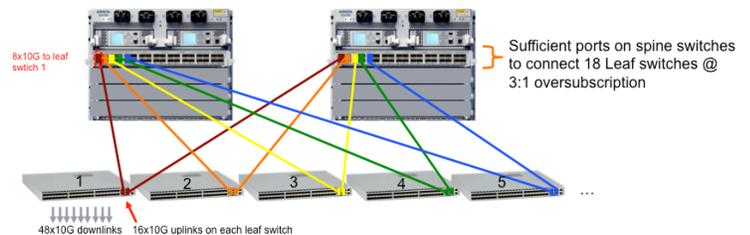


그림 5: 확장형 설계의 초기 구성: 라인 카드가 하나씩 포함된 스위치 1 쌍

리프 스위치가 더 추가되고 스파인 스위치의 첫 번째 라인 카드에 있는 포트가 사용되면 두 번째 라인 카드가 각 새시에 추가되며 링크의 절반이 두 번째 라인 카드로 이동합니다. 그림 6 에 나와 있는 것처럼 이 설계는 리프 스위치가 18 개에서 36 개로 확장될 수 있습니다(엔드 투 엔드 초과 사용 비율 3:1 에서 10G 연결 장치 1,728 개).



그림 6: 확장형 설계의 최초 스파인 확장: 두 번째 라인 카드 모듈 추가

이 프로세스는 여러 번 반복됩니다. 리프와 스파인 간의 업링크 속도가 10G 이면 각 스위치의 라인 카드 4 개에서 포트 4 개에 각 업링크를 분산시킬 수 있습니다.

이 설계의 최종 확장 비율은 스파인 스위치의 포트 확장/밀도, 원하는 초과 사용 비율, 스파인 스위치 수의 함수입니다. 스파인 스위치가 2 개인 경우 레이어 2 또는 레이어 3 에서 해당 설계를 구축할 수 있습니다. Arista 7504E 스파인 스위치 2 개의 최종 확장 비율은 리프 스위치 72 개(엔드 투 엔드 초과 사용 비율 3:1 에서 10G 포트 3,456 개)입니다. Arista 7508E 스위치 한 쌍을 사용하도록 설계하면 스위치 수가 2 배로 증가합니다. 즉, 그림 7 에서처럼 리프 스위치는 144 개가 되고 엔드 투 엔드 초과 사용 비율 3:1 에서 10G 포트 수는 6,912 개가 됩니다.



그림 7: 확장형 설계의 최종 스파인 확장: 각 Arista 7504E 에 네 번째 라인 카드 모듈 추가

25G 또는 50G 업링크

25G 및 50G 브레이크아웃을 지원하는 100G 포트가 포함된 Arista 스위치의 경우 25G 및 50G 브레이크아웃 기능을 통해 100G 포트를 4 x 25G 또는 2 x 50G 로 브레이크아웃할 수 있으므로 레이어 3 ECMP 설계에 대해 보다 범위가 넓은 팬아웃(fan-out)을 사용할 수 있습니다. 이러한 설계를 사용하면 확장형 설계에서 스파인 스위치 수를 늘릴 수 있으므로 팬아웃(fan-out) 범위를 늘려 스파인과 리프 스위치를 더 많이 포함할 수 있습니다.

레이어 2 또는 레이어 3

레이어 2(VLAN 의 모든 위치) 또는 레이어 3(서브넷)에 2 계층 스파인/리프 네트워크를 구축할 수 있습니다. 각 레이어는 장단점이 있습니다.

레이어 2 설계의 경우 유연성이 가장 뛰어나므로 원하는 모든 위치로 VLAN 을 확장하고 MAC 주소를 어떤 위치로든 마이그레이션할 수 있습니다. 단점으로는 하나의 공통적인 결함 도메인(매우 클 수 있음)이 발생하며, 네트워크 내 최소 스위치의 MAC 주소 테이블 크기에 따라 규모 확장이 제한되므로 문제를 해결하기 어려울 수 있다는 단점이 있습니다. 그리고 L3 확장 및 컨퍼전스 시간은 L3 게이트웨이의 호스트 라우트 테이블 크기에 의해 결정되며, 최대 비차단 팬아웃(fan-out) 네트워크는 MLAG(Multi-Chassis Link Aggregation)를 사용하는 스위치 2 개 크기의 스파인 레이어가 됩니다.

레이어 3 설계의 경우 집중 시간이 가장 빠르며, 최대 32 개 이상의 active/active 스파인 스위치를 지원하는 ECMP(Equal Cost Multi Pathing)를 사용하는 팬아웃(fan-out)을 통해 최대 규모로 확장할 수 있습니다. 이러한 설계는 L2/L3 게이트웨이를 첫 번째 홉 스위치로 지역화하므로 스위치 간에 서로 느려지게 만드는 현상(최소 공통분모 적용) 없이 다양한 스위치 클래스를 최대 용량까지 활용할 수 있는 뛰어난 유연성이 제공됩니다.

하지만 레이어 3 설계에서는 VLAN 및 MAC 주소를 단일 스위치 또는 스위치 쌍으로만 이동하게 제한되어 있으므로 따라서 VM 이동성의 범위도 단일 스위치 또는 스위치 쌍 범위로 제한됩니다. 이는 일반적으로 하나의 랙이나 최대 몇 개의 랙 범위입니다.

VXLAN 오버레이가 포함된 레이어 3 언더레이

VXLAN 은 비전매 멀티 벤더 VXLAN 표준을 통해 레이어 3 언더레이 전체에서 레이어 2 오버레이를 사용할 수 있도록 함으로써 레이어 3 설계를 보완합니다. VXLAN 에서는 레이어 3 설계의 장점(확장, 대규모 네트워크 범위, 빠른 집중 시간, 결함 도메인 최소화)과 레이어 2의 유동적인 특성(VLAN 및 MAC 주소 이동성)을 함께 제공하므로 레이어 2 와 레이어 3 설계의 단점을 모두 방지할 수 있습니다.

가상 서버 인프라의 일부분으로 가상 스위치를 통해 소프트웨어에서 VXLAN 기능을 사용할 수 있습니다. 이러한 방식에서는 레이어 2 가 레이어 3 의 범위를 넘어 확장되지만, 트래픽이 최적의 방식으로 올바른 물리적 서버에 도달하도록 하지는 않습니다. 또한 네트워크에서 VXLAN 또는 기타 오버레이를 배치하는 소프트웨어 기반 방식을 사용하는 경우 NIC 의 오프로드 기능이 사용되지 않으므로 서버의 CPU 주기가 증가합니다.

Arista 스위치의 하드웨어 VXLAN 게이트웨이 기능을 사용하면 유동성과 확장성을 최대화하고 트래픽을 최적화할 수 있습니다. 물리적 네트워크는 레이어 3 에 유지되므로 최대 규모로 확장이 가능하고, 테이블/기능을 가장 적절하게 활용할 수 있으며, 집중 시간을 최고 레벨로 높일 수 있습니다. 서버는 NIC CPU 오프로드 기능을 계속 제공하며 VXLAN 하드웨어 게이트웨이는 레이어 3 포워딩에 대한 레이어 2 오버레이 기능과 함께 레이어 2 및 레이어 3 포워딩 기능을 제공합니다.

표 4: 레이어 2, 레이어 3, VXLAN 설계가 적용된 레이어 3의 장단점

설계 유형	장점	단점
레이어 2	모든 위치로 VLAN을 확장 가능하므로 뛰어난 유연성 제공 MAC를 이동할 수 있으므로 원활한 VM 이동성	단일(대규모) 결합 도메인 STP로 인한 중복/HA 링크 차단 장애 도메인을 확장하지 않고서는 포트 또는 데이터센터의 범위를 넘어 확장하기 어려움 컨트롤 플레인의 속도(초당 ARP)로 인해 L3 게이트웨이 집중성이 떨어짐 L3 확장 범위가 L3 게이트웨이의 호스트 라우트 확장 범위에 의해 결정됨 최대 2방향으로만 확장이 가능함 (MLAG active/active) 최대 VLAN x 스위치의 포트 수가 스페닝 트리 논리적 포트 수 범위에 따라 제한됨 문제 해결이 어려움
레이어 3	여러 포트 또는 데이터센터로 확장됨 ECMP를 통해 초대형 확장 가능 집중/재집중 시간이 매우 빠름	VLAN이 단일 스위치로 제약됨 단일 스위치 내에서만 MAC 이동성
VXLAN 오버레이가 포함된 레이어 3 언더레이	VXLAN을 통해 모든 스위치/장치로 VLAN 확장 가능 L3 연결이 가능한 모든 위치로 MAC 이동성 여러 포트 또는 데이터센터로 확장됨 MAC를 이동할 수 있으므로 원활한 VM 이동성 ECMP를 통해 초대형 확장 가능 집중/재집중 시간이 매우 빠름	VXLAN 기반의 소프트웨어/하이퍼바이저 가상 스위치가 호스트에 CPU 오버헤드를 발생시킴 (하드웨어 VXLAN 게이트웨이에는 이러한 특성이 없음)

하드웨어 VXLAN 게이트웨이 기능이 있는 Arista 스위치 플랫폼에는 'E' 또는 'X'자가 붙은 모든 Arista 스위치(Arista 7500E Series, Arista 7280E, Arista 7320X Series, Arista 7300E Series, Arista 7060X Series, Arista 7050X Series)와 Arista 7150S Series가 포함됩니다.

이러한 플랫폼은 Arista CloudVision, OVSDb와 같은 개방형 표준 기반의 비전매 프로토콜 또는 정적 구성을 통한 통합을 수행함으로써 유니캐스트 기반의 하드웨어 VXLAN 게이트웨이 기능을 지원합니다. 이처럼 하드웨어 VXLAN 게이트웨이 기능이 개방적으로 제공되므로 최종 사용자는 고객이 벤더에 고착화되지 않고도 적절한 클라우드 통합 플랫폼을 선택할 수 있습니다.

포워딩 테이블 크기

이더넷 스위칭 ASIC는 포워딩을 결정하기 위해 MAC 테이블(L2), 호스트 라우트 테이블(L3), L3 프리픽스 룩업을 위한 LPM(LONGEST PREFIX MATCH) 등의 여러 포워딩 테이블을 사용합니다. L2 또는 L3에 구축할 수 있는 최대 네트워크 크기는 이러한 테이블의 크기에 따라 결정됩니다.

기존에는 서버나 호스트의 MAC 주소와 IP 주소가 각각 하나씩만 있었습니다. 서버 가상화를 사용하면 가상 서버당 MAC 주소와 IP 주소를 최소 하나씩 포함할 수 있으며, vNIC(가상 NIC)를 추가로 정의하면 주소/VM을 추가로 얻을 수 있습니다. 대부분의 IT 조직은 듀얼 IPv4/IPv6 스택을 배치하고 있거나 향후 배치할 예정입니다. 따라서 스위치의 포워딩 테이블은 IPv4 및 IPv6 테이블 요구 사항을 모두 고려해야 합니다.

네트워크를 레이어 2 에 구축하는 경우 모든 스위치가 네트워크에 있는 MAC 주소를 전부 인식하게 됩니다. 그리고 스파인에 있는 스위치는 레이어 2 와 레이어 3 간의 포워딩 기능을 제공하며, 게이트웨이 호스트 라우트를 제공해야 합니다.

네트워크를 레이어 3 에 구축하는 경우 스파인 스위치는 리프 스위치당 서브넷 1~2 개에 대해서만 IP 포워딩을 사용하면 되며 호스트 MAC 주소는 몰라도 됩니다. 리프 스위치는 해당 스위치에 로컬인 IP 호스트 라우트 및 MAC 주소를 알아야 하지만 로컬 연결 외부의 라우트와 주소는 확인할 필요가 없습니다. 이 경우 리프 스위치에 필요한 라우팅 프리픽스는 스파인 스위치로 이동하는 단일 디폴드 라우트뿐입니다.

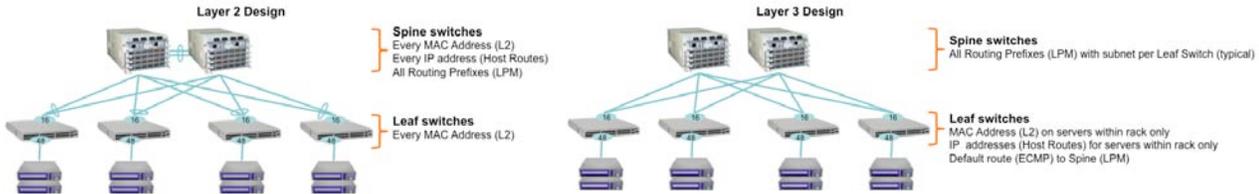


그림 8: 레이어 2 및 레이어 3 설계 비교

네트워크가 구축되는 위치(레이어 2 또는 레이어 3)에 관계없이 네트워킹 테이블 크기는 VM 수에 좌우되는 경우가 많습니다. 현재 최신 x86 서버는 CPU 코어/소켓 6 개 또는 8 개의 듀얼 소켓입니다. 일반적인 기업 작업 부하는 CPU 당 10 개의 VM 코어를 허용하므로 일반적인 서버에서 VM 60~80 개가 실행되는 경우도 드물지 않습니다. 이 VM 수는 향후 계속해서 늘어만 갈 것입니다.

CPU 당 10 개의 VM 그리고 쿼드 코어 CPU 를 사용하며 서버당 소켓은 2 개, 랙당 물리적 서버는 40 대, 서버 랙은 20 개인 설계에서 네트워크의 포워딩 테이블에 대한 요구 사항은 다음과 같습니다.

표 5: 레이어 2 및 레이어 3 설계의 포워딩 테이블 확장 특성

포워딩 테이블	레이어 2 스파인/리프 설계		레이어 3 스파인/리프 설계	
	스�파인 스위치	리프 스위치	스�파인 스위치	리프 스위치
MAC 주소 (VM 당 vNIC 1 개)	VM 당 MAC 주소 1 개 x CPU 당 VM 10 개 x 소켓당 CPU 4 개 x 서버당 소켓 2 개 = 서버당 VM 80 개 x 랙당 서버 40 대 = 랙당 MAC 주소 3,200 개 x 랙 20 개 = MAC 주소 64,000 개		최소 (스�파인 스위치는 L3 에서 작동하므로 L2 포워딩 테이블은 사용되지 않음)	VM 당 MAC 주소 1 개 x CPU 당 VM 10 개 x 소켓당 CPU 4 개 x 소켓 2 개 = 서버당 VM 80 개 x 랙당 서버 40 대 = MAC 주소 3,200 개
IP 라우트 LPM	소수의 IP 접두사	없음 (리프 스위치가 L2 에서 작동하며 L3 은 없음)	랙당 서브넷 1 개 x 랙 20 개 = IP 라우트 LPM 접두사 20 개	최소 (스�파인 스위치로의 단일 ECMP 라우트)
IP 호스트 라우트 (IPv4 전용)	VM 당 IPv4 호스트 라우트 1 개 랙당 IPv4 호스트 라우트 3,200 개 x 랙 20 개 = IP 호스트 라우트 64,000 개	없음 (리프 스위치가 L2 에서 작동하며 L3 은 없음)	최소 (스�파인 스위치에 IP 호스트 라우트 없음)	VM 당 IPv4 호스트 라우트 1 개 랙당 IPv4 호스트 라우트 3,200 개 = IP 호스트 라우트 3,200 개
IP 호스트 라우트 (IPv4 + IPv6 듀얼 스택)	VM 당 IPv4 및 IPv6 호스트 라우트 1 개 IPv4 호스트 라우트 64,000 개 + IPv6 호스트 라우트 64,000 개	없음 (리프 스위치가 L2 에서 작동하며 L3 은 없음)	최소 (스�파인 스위치에 IP 호스트 라우트 없음)	VM 당 IPv4 및 IPv6 호스트 라우트 1 개 IPv4 호스트 라우트 3,200 개 + IPv6 호스트 라우트 3,200 개

레이어 2 스페닝 트리 논리적 포트 수 확장

대형 레이어 2 네트워크에서 일반적으로 발생하는 문제(대규모 브로드캐스트 도메인, 단일 결함 도메인, 문제 해결의 어려움) 외에도 간과하기 쉬운 한 가지 제한 요인은 스위치에서 스페닝 트리 프로토콜을 실행하는 경우 발생하는 컨트롤 플레인 CPU 오버헤드입니다. 프로토콜로서 스페닝 트리는 프로토콜에 장애가 발생하는 경우 최근 흔히 보이는 '결함 시 닫힘' 상태가 아닌 '결함 시 열림' 상태로 설정된다는 점에서 상대적으로 독특하다고 할 수 있습니다. 어떤 이유로든 프로토콜 장애가 발생하면 네트워크 루프가 나타납니다. 이와 같은 스페닝 트리의 특성을 고려할 때 스위치 컨트롤 플레인에 과도한 부담을 주어서는 안 됩니다.

스witch는 RPVST(Rapid Per VLAN Spanning Tree)를 사용하여 각 VLAN에 대해 스페닝 트리의 여러 독립적 인스턴스를 유지하여 일정한 간격으로 포트에서 BPDU를 보내고 받으며 해당 BPDU를 기준으로 물리적 포트의 포트 상태를 학습/수신/포워딩/차단으로 변경합니다. VLAN 트렁킹을 면밀하게 설계하지 않으면 동기화되지 않는 독립 인스턴스를 여러 개 관리하는 경우 확장 관련 문제가 발생할 수 있습니다. 예를 들어 단일 포트에서 VLAN 4,000개를 트렁킹하는 경우 각 VLAN의 상태를 개별적으로 추적해야 합니다.

따라서 RPVST보다는 MSTP(다중 스페닝 트리 프로토콜)를 사용하는 것이 좋습니다. 그 이유는 작동하는 스페닝 트리 프로토콜 인스턴스 수가 더 적으며, 상태 간에 물리적 포트를 이동하는 작업은 그룹으로 수행될 수 있기 때문입니다. 그러나 이와 같은 개선 사항에도 불구하고 레이어 2 논리적 포트 수는 계속해서 적절하게 관리되어야 합니다.

스패닝 트리에 포함되는 스위치의 개별 확장 특성은 다양하지만 설계 시 넣어야 할 주요 요인은 다음과 같습니다.

- 지정한 스위치에서 지원되는 STP 논리적 포트의 수 (VlanPorts의 수라고도 함)
- RPVST를 사용 중인 경우 지원되는 스페닝 트리의 인스턴스 수

스패닝 트리를 사용한 대형 레이어 2 설계에 걸친 레이어 3 위로 레이어 2를 연장하기 위한 오버레이를 제공하는 데 사용되는, VXLAN을 활용한 레이어 3 ECMP 설계를 항상 사용하는 것이 좋습니다. 레이어 3과 VXLAN이 포함된 설계를 사용하는 경우 유동성과 확장성을 최대화하고 트래픽을 최적화할 수 있을 뿐 아니라 장애 도메인을 최소화하며 테이블/용량을 가장 적절하게 활용하고 집중성을 최고로 높일 수 있습니다.

ARISTA 2 계층 스파인/리프 확장형 설계

2 계층 스파인/리프 설계에서는 모든 리프 스위치가 모든 스파인 스위치에 연결됩니다. 이 설계는 레이어 2 나 레이어 3 에 구축할 수 있지만 레이어 3 설계의 확장성이 더 뛰어납니다. 레이어 3 설계는 스파인 스위치를 3 개 이상 포함할 수 있으며 MAC 엔트리 및 호스트 라우트가 지정 리프 스위치 또는 리프 스위치 쌍으로 지역화되기 때문입니다.

ARISTA 7500 SERIES(스파인 2 개)를 사용하는 스파인/리프 설계 1G 노드

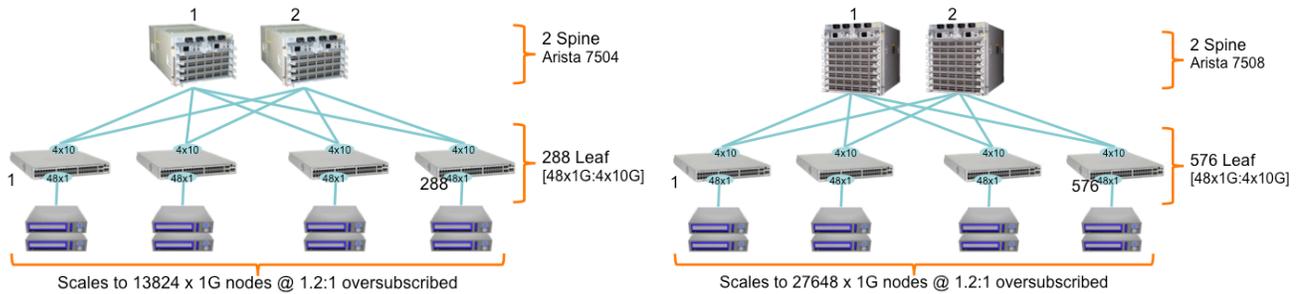


그림 9: 10G 업링크에서 Arista 7504E/7508E 스파인 스위치(최대 2 개 스위치까지 확장 가능)를 사용하는 1G 연결 노드의 스파인/리프 네트워크 설계

ARISTA 7500 SERIES(스파인 2 개)를 사용하는 스파인/리프 설계 10G 노드(초과 사용 비율 3:1)

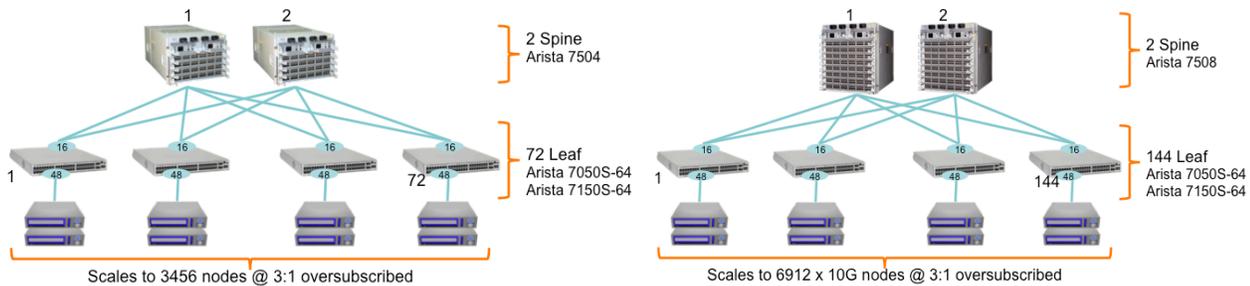


그림 10: 10G 업링크에서 Arista 7504E/7508E 스파인 스위치(최대 2 개 스위치까지 확장 가능)를 사용하는 10G 연결 노드의 스파인/리프 네트워크 설계(초과 사용 비율 3:1)

ARISTA 7500 SERIES(스파인 2 개)를 사용하며 초과 사용되지 않는 스파인/리프 설계 10G 노드

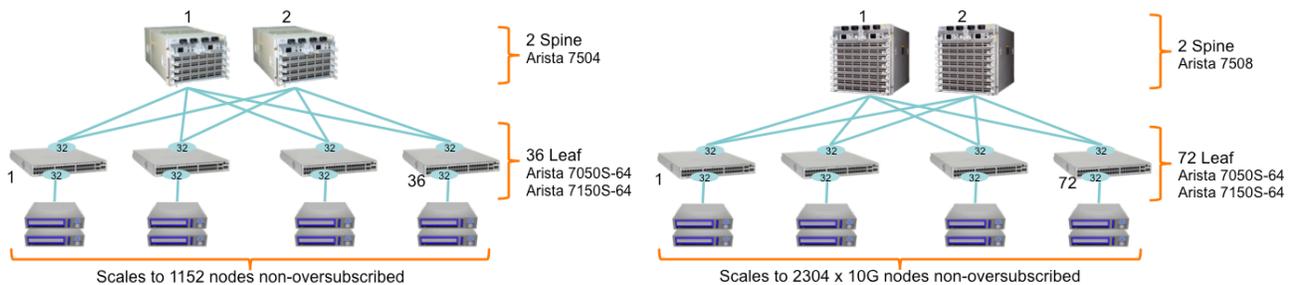


그림 11: 10G 업링크에서 Arista 7504E/7508E 스파인 스위치(최대 2 개 스위치까지 확장 가능)를 사용하며, 초과 사용되지 않는 10G 연결 노드의 스파인/리프 네트워크 설계

이러한 토폴로지는 모두 레이어 2 또는 레이어 3 에 구축할 수 있습니다. 레이어 2 설계의 경우 MLAG 가 active/active 로 실행되며 차단된 링크가 없는 L2 네트워크를 제공합니다. 이러한 네트워크에서는 스파인 스위치 간에 MLAG 피어 링크를 사용해야 합니다.

리프 스위치에서 MLAG 를 사용하여 active/active 방식으로 서버/스토리지를 연결할 수도 있습니다. 이 경우 리프 스위치 쌍은 MLAG 쌍이 되며 스위치 간에 MLAG 피어 링크가 설정됩니다. MLAG 는 듀얼 연결 장치용 스위치에 대해 로컬 상태로 유지되도록 네트워크 트래픽을 우선적으로 처리하므로 MLAG 피어 링크는 비교적 적은 수(2 개 이상)의 물리적 링크일 수 있습니다.

10G 업링크를 사용하는 대규모 스파인/리프 설계

ECMP 레이아웃에서 스파인 스위치를 최대 128 개까지 포함한 레이어 3 을 사용해 설계를 확장할 수도 있으며, 이 경우 리프 스위치의 초대형 팬아웃이 가능합니다. 시간이 지나면서 네트워크가 확장됨에 따라 스파인 스위치에 라인 카드 모듈이 추가되는 그런 방식을 스파인 스위치에도 사용할 수 있습니다. 네트워크의 스파인 스위치 수를 2 개에서 4 개, 8 개, 16 개로 늘릴 수 있으며 최대 128 개로 늘릴 수 있습니다. 스파인과 리프 간의 모든 경로는 BGP, OSPF 등의 표준 라우팅 프로토콜을 사용하여 active/active 모드로 실행되며 최대 128 방향의 ECMP 를 사용하여 모든 경로를 active/active 모드로 실행할 수 있습니다. 아래 그림 12~14 에서는 초과 사용 비율이 3:1 인 10G 설계에서 네트워크의 스파인 스위치 수를 4 개에서 8 개와 16 개로 늘리는 방법을 보여 줍니다.

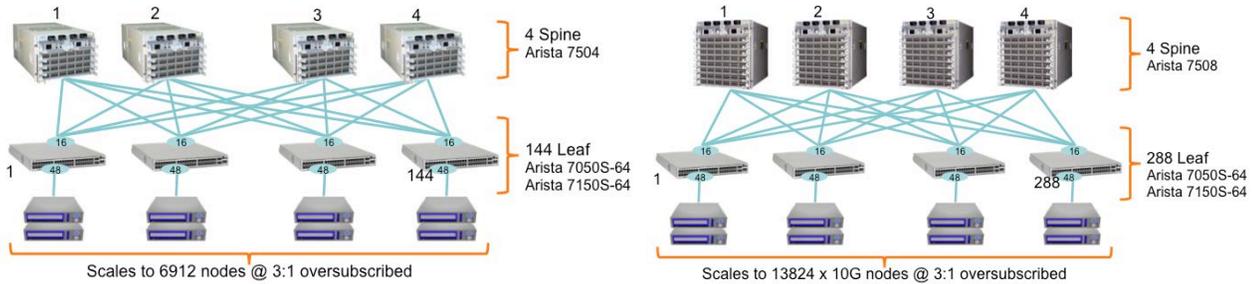


그림 12: Arista 7504E/7508E 스파인 4 방향 ECMP 에서 Arista 64 포트 10G 리프 스위치로 확장(초과 사용 비율 3:1)

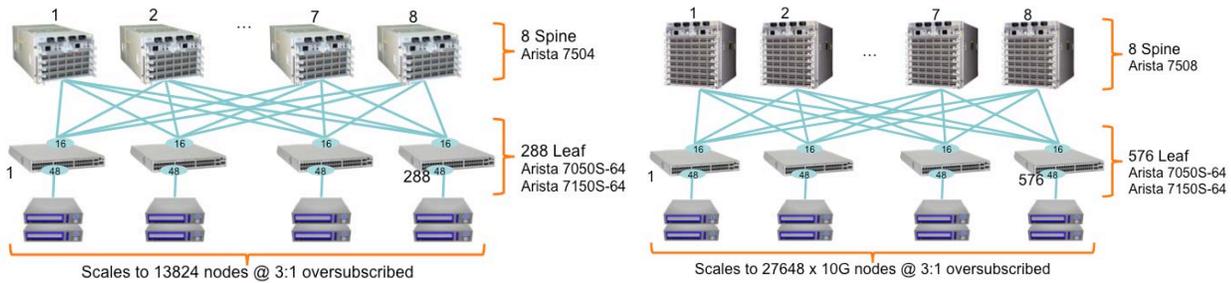


그림 13: Arista 7504E/7508E 스파인 8 방향 ECMP 에서 Arista 64 포트 10G 리프 스위치로 확장(초과 사용 비율 3:1)

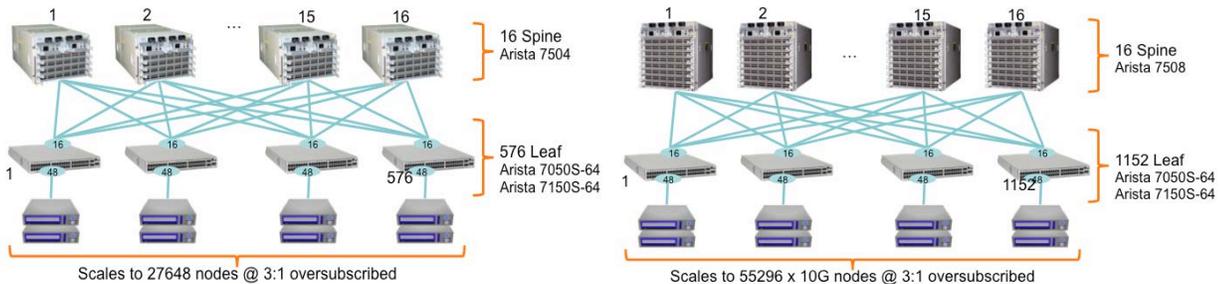


그림 14: Arista 7504E/7508E 스파인 16 방향 ECMP 에서 Arista 64 포트 10G 리프 스위치로 확장(초과 사용 비율 3:1)

다음 그림에서는 4 방향 ECMP 를 통해 1G 서버 설계를 확장하는 방법을 보여 줍니다. 각 리프 스위치에는 48x1G 서버/스토리지 연결용 4x10G 업링크가 있습니다.

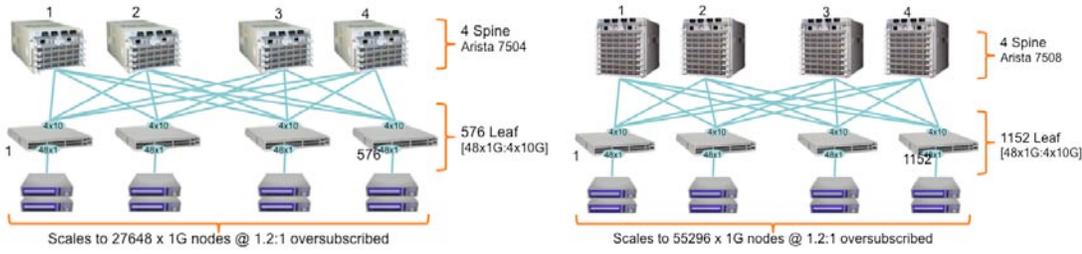


그림 15: Arista 7504E/7508E 스파인 4 방향 ECMP 에서 Arista 48x10G 리프 스위치로 확장(초과 사용 비율 1.2:1)

이와 동일한 디자인 원칙을 적용하여 초과 사용되지 않는 10G 네트워크를 구축할 수 있습니다. 종량제 방식의 요금제로 초반에는 비교적 저렴한 자본적 지출로 투자하다가 차후에 네트워크 크기를 점점 늘릴 수 있습니다.

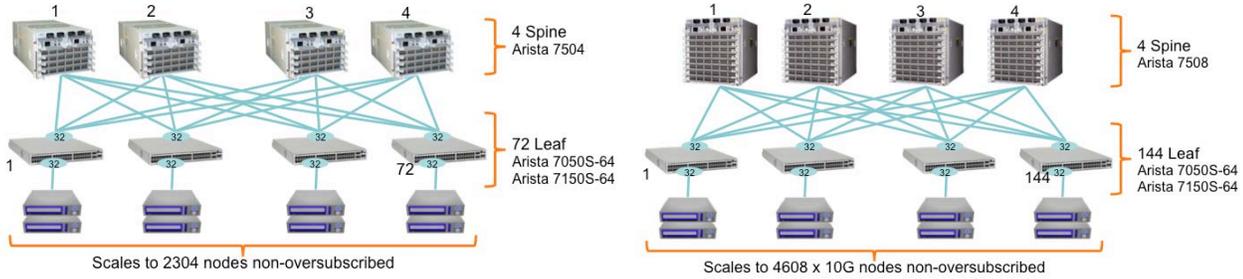


그림 16: Arista 7504E/7508E 스파인 4 방향 ECMP 에서 Arista 64 포트 10G 리프 스위치로 확장(초과 사용되지 않음)

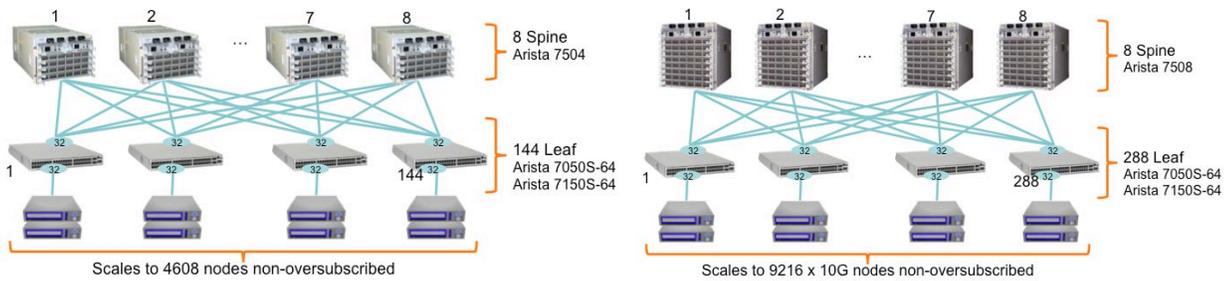


그림 17: Arista 7504E/7508E 스파인 8 방향 ECMP 에서 Arista 64 포트 10G 리프 스위치로 확장(초과 사용되지 않음)

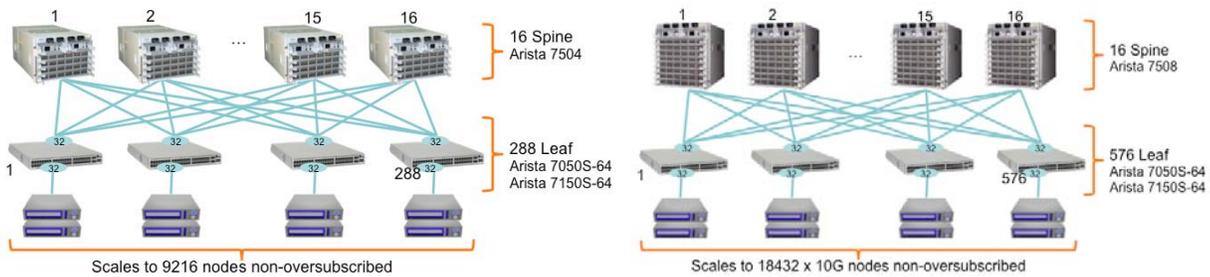


그림 18: Arista 7504E/7508E 스파인 16 방향 ECMP 에서 Arista 64 포트 10G 리프 스위치로 확장(초과 사용되지 않음)

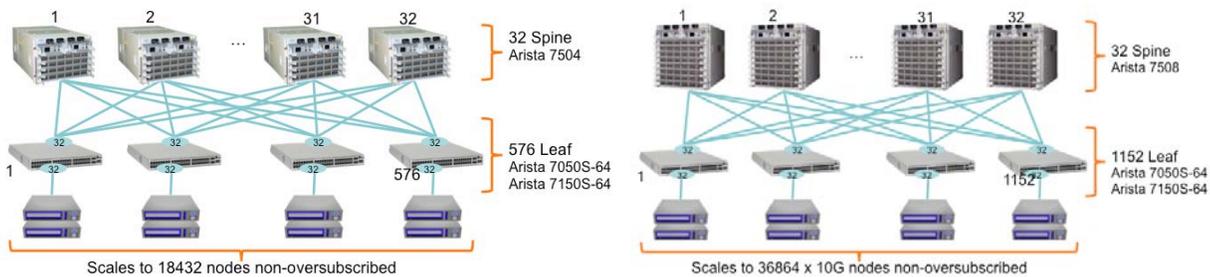


그림 19: Arista 7504E/7508E 스파인 32 방향 ECMP 에서 Arista 64 포트 10G 리프 스위치로 확장(초과 사용되지 않음)

40G 업링크를 사용하는 대규모 설계

위의 그림과 같은 단순한 디자인 원칙을 사용하여 스파인/리프 간에 10G 가 아닌 40G 업링크를 사용한 네트워크를 구축할 수 있습니다. Arista 스위치에서 40G QSFP+ 포트는 40G 1 개 또는 10G 4 개로 구성할 수 있으며, 광 기능을 사용하면 개별 10G 링크로 브레이크아웃할 수 있습니다. 대부분의 설계는 10G 업링크에서 40G 업링크로 쉽게 확장 가능하거나 10G 와 40G 조합을 지원합니다. AgilePorts 를 지원하는 Arista 7150S 등의 Arista 스위치 플랫폼에서는 SFP+ 인터페이스 4 개를 40G 포트로 구성하여 업링크 속도 조합을 보다 유연적으로 선택할 수 있습니다.

아래 다이어그램에서는 초과 사용 비율이 3:1 인 10G 노드용 레이어 3 ECMP 설계에서 리프에서 스파인으로의 40G 업링크를 사용한 최대 확장 범위를 보여 줍니다.

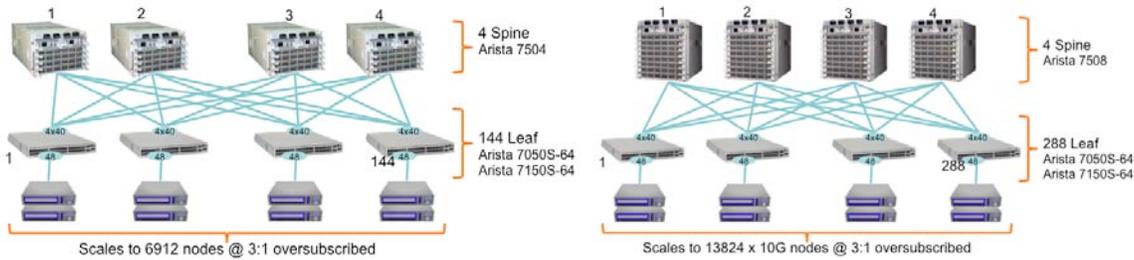


그림 20: Arista 7504E/7508E 스파인 4 방향 ECMP 에서 Arista 리프 스위치 48x10G + 4x40G 로 확장(초과 사용 비율 3:1)

옵틱, 케이블 그리고 트랜시버 선택사항들

다양한 트랜시버, 광학 기능, 케이블 중에서 선택할 수 있습니다. **SFP+/SFP** 는 가장 널리 사용되며 폭넓은 거리를 지원하는 10G/1G 용 트랜시버입니다.

표 6: SFP+/SFP 트랜시버 옵션

유형	속도	범위	매체	참고 사항
10GBASE-CR	10G	0.5m, 1m, 1.5m, 2m, 2.5m, 3m, 5m, 7m	DAC(직접 연결) CX1 Twinax	단자 사전 처리(양 끝의 트랜시버가 카퍼 케이블에 퓨즈로 연결됨)
10GBASE-SRL	10G	100m(OM3), 150m(OM4)	50μ MMF	최대 100m 까지 10GBASE-SR 과 광학 방식으로 상호 운용 가능
10GBASE-SR	10G	100m(OM3), 150m(OM4)	50μ MMF	최대 100m 까지 10GBASE-SRL 과 광학 방식으로 상호 운용 가능
10GBASE-AOC	10G	3m, 5m, 7m, 10m, 15m, 20m, 25m, 30m	단자 사전 처리된 광 트랜시버	단자 사전 처리(양 끝의 트랜시버가 광 케이블에 퓨즈로 연결됨)
10GBASE-LRL	10G	1km	9μ SMF	최대 1km 까지 10GBASE-LR 과 광학 방식으로 상호 운용 가능
10GBASE-LR	10G	10km	9μ SMF	최대 1km 까지 10GBASE-LRL 과 광학 방식으로 상호 운용 가능
10GBASE-ER	10G	40km	9μ SMF	
10GBASE-ZR	10G	80km	9μ SMF	
10GBASE-DWDM	10G	40km/80km	9μ SMF	43 개의 파장 사용 가능
1000BASE-T	100M/1G	100m	Cat5e	
1000BASE-SX	1G	550m	50μ MMF	
1000BASE-LX	1G	10km	9μ SMF	

QSFP+ 트랜시버는 40G 연결에 사용되며, 단일 물리적 포트를 40G 1 개 또는 10G 4 개로 브레이크아웃하는 데에도 사용됩니다.



표 7: QSFP+ 트랜시버 옵션

유형	속도	범위	매체	참고 사항
40GBASE-CR4	40G	0.5m, 1m, 2m, 3m, 5m	DAC(직접 연결)	단자 사전 처리(양 끝의 트랜시버가 카퍼 케이블에 퓨즈로 연결됨)
40GBASE-CR4~10GBASE-CR 4 개	10G 4 개	0.5m, 1m, 2m, 3m, 5m	DAC(직접 연결)	단자 사전 처리(한쪽 끝에는 QSFP+, 반대쪽 끝에는 SFP+ 4 개가 카퍼 케이블에 퓨즈로 연결됨)
40GBASE-AOC	40G	3m, 5m, 7m, 10m, 15m, 20m, 25m, 30m, 50m, 75m, 100m	AOC(활성 광 매체)	단자 사전 처리(양 끝의 트랜시버가 광 케이블에 퓨즈로 연결됨)
40GBASE-SR4	40G	100m(OM3), 150m(OM4)	50μ MMF MTP12	40G 1 개(40GBASE-SR4) 또는 10G 4 개(10GBASE-SR/SRL 과 호환됨)로 작동 가능
40GBASE-XSR4	40G	300m(OM3), 400m(OM4)	50μ MMF MTP12	40G 1 개(40GBASE-XSR4) 또는 10G 4 개(10GBASE-SR/SRL 과 호환됨)로 작동 가능 최대 150m 까지 40GBASE-SR4 와 호환됨
40GBASE-UNIV	40G	150m(50μ MMF) 500m(9μ SMF)	50μ MMF 또는 9μ SMF LC	MMF 또는 SMF 로 작동함 최대 500m 까지 SMF 를 통해 40GBASE-LR4 와 호환됨
40GBASE-SRBD	40G	100m(OM3), 150m(OM4)	50μ MMF LC	40G BiDi 라고도 함 기존 MMF LC 파이버 플랜트를 통해 40G 에 40G 연결 가능
40GBASE-LRL4	40G	1km	9μ SMF LC	최대 1km 까지 저렴한 비용으로 이중 파이버 사용 가능 최대 1km 까지 40GBASE-LR4 와 호환됨
40GBASE-PLRL4	40G	1km	9μ SMF MTP12	40G 1 개(40G-PLRL4) 또는 10G 4 개(최대 1km 까지 10GBASE-LR/LRL 과 호환됨)로 작동 가능
40GBASE-PLR4	40G	10km	9μ SMF MTP12	40G 1 개(40G-PLR4) 또는 10G 4 개(최대 10km 까지 10GBASE-LR 과 호환됨)로 작동 가능
40GBASE-LR4	40G	10km	9μ SMF LC	모든 40GBASE-LR4 광학 기능과 멀티 밴더 상호 운용 가능/최대 1km 까지 40G-LRL4 와 상호 운용 가능
40GBASE-ER4	40G	40km	9μ SMF LC	

임베디드 100G 광학 기능은 Arista 7500E-72S-LC 및 7500E-12CM-LC 라인 카드 모듈에 사용되며 트랜시버 없이도 업계 표준 100GBASE-SR10, 40GBASE-SR4, 10GBASE-SR 포트를 제공합니다. 이를 통해 비용 효율성과 밀도가 업계 최고 레벨인 10/40/100G 연결이 제공됩니다. 각 포트는 표준 MPO/MTP24 케이블과 연결되며(케이블/커넥터 하나에 파이버 쌍 12 개) 구성의 유연성이 매우 뛰어나므로 포트 하나를 다음 중 원하는 구성으로 작동할 수 있습니다.

- 100GBASE-SR10 1 개
- 40GBASE-SR4 3 개
- 40GBASE-SR4 2 개 + 10GBASE-SR 4 개
- 40GBASE-SR4 1 개 + 10GBASE-SR 8 개
- 10GBASE-SR 12 개



이러한 포트를 OM3/OM4 MMF 와 함께 사용하면

300m(OM3)와 400m(OM4) 거리를 지원할 수 있습니다. LC 패치 케이블 12 개에 MPO 를 사용하여 LC 커넥터 12 개로 브레이크아웃하면 SFP+에 연결할 수 있습니다.

일부 스위치에서 제공되는 **Arista AgilePorts** 는 SFP+ 포트 4 개 또는 10 개로 구성된 그룹을 사용하여 업계 표준 40GBASE-SR4 또는 100GBASE-SR10 을 생성할 수 있습니다. 그러면 네트워크를 10G 에서 40G, 100G 로 더욱 유연하게 확장 및 개선함과 동시에 거리도 더욱 유동적으로 지원할 수 있습니다.

QSFP100 트랜시버는 100G 데이터센터 범위 연결용으로 최적화되어 있습니다. 일부 스위치 플랫폼에는 25G 4 개/50G 2 개로 브레이크아웃하는 옵션도 있습니다.

표 8: QSFP100 트랜시버 옵션

유형	속도	범위	매체	참고 사항
100GBASE-CR4	100G	1m, 2m, 3m, 5m	DAC(직접 연결)	단자 사전 처리(양 끝의 트랜시버가 카퍼 케이블에 퓨즈로 연결됨)
100G~25G 4 개	25G 4 개	0.5m, 1m, 2m, 3m, 5m	DAC(직접 연결)	단자 사전 처리(한쪽 끝에는 QSFP100, 반대쪽 끝에는 QSFP+ 4 개가 카퍼 케이블에 퓨즈로 연결됨)
100GBASE-AOC	100G	3m, 5m, 7m, 10m, 15m, 20m, 25m, 30m	AOC(활성 광 매체)	단자 사전 처리(양 끝의 트랜시버가 광 케이블에 퓨즈로 연결됨)
100GBASE-SR4	100G	70m(OM3) 100m(OM4)	50μ MMF MTP12	일부 스위치 플랫폼에서는 100G 1 개 또는 25G 4 개로 작동 가능함
100GBASE-LRL4	100G	1km	9μ SMF LC	최대 1km 까지 저렴한 비용으로 이중 파이버 사용 가능 최대 1km 까지 100GBASE-LR4 와 호환됨
100GBASE-LR4	100G	10km	9μ SMF LC	모든 100GBASE-LR4 광학 기능과 멀티 번더 상호 운용 가능/최대 1km 까지 100G-LRL4 와 상호 운용 가능

CFP2 트랜시버는 동일한 대도시 지역 내 데이터센터 간의 장거리 연결용으로 최적화되어 있습니다.

표 9: CFP2 트랜시버 옵션

유형	속도	범위	매체	참고 사항
100GBASE-XSR10	100G	300m(OM3), 400m(OM4)	50μ MMF MTP24	100GBASE-SR10 과 호환됨 최대 100m(OM3)/150m(OM4)
100GBASE-LR4	100G	10km	9μ SMF LC	
100GBASE-ER4	100G	40km	9μ SMF LC	

이러한 설계를 적용할 수 있도록 하는 ARISTA EOS 의 기반이 되는 기능

개방형 아키텍처

Arista의 솔루션을 사용하는 고객은 특정 벤더의 제품이나 전매 프로토콜에 고착화되는 일 없이 원하는 아키텍처를 자유롭게 선택할 수 있습니다. Arista UCN(범용 클라우드 네트워크) 아키텍처는 전적으로 BGP, OSPF, LACP, VXLAN 등의 표준 기반 도구를 토대로 하므로 고객은 자신의 필요에 따라 제품을 완벽하게 상호 운용할 수 있습니다. 이러한 방식은 향후 네트워킹 요구 사항이 변경되는 경우 고객이 제품을 유동적으로 조정할 수 있게 합니다.

모든 레이어에서의 프로그래밍 기능

대부분의 네트워킹 벤더는 자사의 시스템이 개방적이며 프로그래밍이 가능하다고 주장하지만, 자세히 살펴보면 프로그래밍 기능이 제한적이며 불완전하게 구현되어 스택의 일부분만을 처리하며 사용하고 유지하기가 어려움을 알 수 있습니다. 고객이 완전하게 제어할 수 있는 운영 체제를 제공한다는 핵심 원칙에 따라 Arista EOS에서는 효율적으로 구조화된 개방형 API를 통해 소프트웨어 스택의 모든 레이어(커널, 하드웨어, 컨트롤 플레인, 관리)에 액세스할 수 있습니다.

이와 같은 혁신적 구조를 통해 EOS는 다음과 같은 최고의 프로그래밍 방식 액세스 기능을 제공합니다.

- Arista EOS SDK - 업계 최초의 프로그래밍/확장 가능 네트워크 플랫폼과 함께 기존에 해당 플랫폼을 배치한 고객의 사례를 제시합니다.
- EOS API - HTTP/HTTPS API를 통해 JSON을 사용하며, 기본 시스템에 완벽하게 액세스할 수 있는 기능을 제공합니다.
- Arista의 개방형 API를 통해 Splunk, F5, Palo Alto, Nuage, VMware 등의 기술 협력사 제품과 통합이 가능합니다.
- Arista CloudVision - Arista는 실시간 모니터링과 멀티 벤더와의 통합을 위한 단일 API로 네트워크 전체에 사용 가능한 다기능 컨트롤러를 제공함으로써 EOS 플랫폼을 확장합니다. CloudVision은 중앙 Sysdb 인스턴스를 제공함으로써 네트워크 내의 모든 스위치 상태를 추상화합니다. 따라서 안정적이며 확장 가능한 방식으로 네트워크 전체를 파악할 수 있습니다.

Arista의 범용 클라우드 네트워크 설계는 수상 경력에 빛나는 Arista 확장 가능 운영 체제의 다음과 같은 여러 기반 기능에 사용되고 있습니다.

MLAG(MULTI-CHASSIS LINK AGGREGATION)

MLAG를 사용하는 장치를 Arista 스위치 쌍(MLAG 쌍)에 연결할 수 있으며, 이 경우 모든 링크는 active/active 모드로 실행됩니다. MLAG에서는 병목 현상이 발생하지 않고, 복구 기능이 제공되며, 레이어 2 링크를 active/active 모드로 작동할 수 있습니다. 또한 STP 차단형 링크에서처럼 대역폭의 50%가 낭비되지 않습니다. MLAG를 사용하는 L3 Anycast Gateway(가상 ARP/VARP)에서는 L3 게이트웨이가 active/active 모드로 작동할 수 있으며 HSRP 또는 VRRP와 같은 프로토콜의 오버헤드가 발생하지 않습니다.

MLAG는 인접 장치에 대해 표준 링크 애그리게이션(LAG)과 같은 방식으로 작동하며, 이전의 IEEE 802.3ad(최근의 IEEE 802.1AX-2008)에 해당하는 LACP(링크 애그리게이션 제어 프로토콜) 또는 정적 '모드 사용' 구성으로 실행될 수 있습니다.

MLAG 스위치 쌍은 스위치 간의 포워딩 상태를 동기화하므로 특정 노드에서 장애가 발생해도 작업이 중단되거나 장애가 발생하지 않습니다. 그 이유는 장치가 active/active 모드로 작동하므로 학습할 수 있는 새로운 상태가 없거나 standby 상태에서 active로 변경될 프로토콜이 없기 때문입니다.

ZTP(ZERO TOUCH PROVISIONING)

ZTP를 사용하는 스위치는 구성을 하지 않아도 물리적으로 배치될 수 있습니다. ZTP를 사용하는 경우 스위치는 네트워크 내의 중앙 위치에서 해당 이미지와 구성을 로드합니다. 따라서 배치가 간소화되므로 스위치를 프로비저닝하거나 네트워크 엔지니어가 직렬 콘솔 케이블을 직접 확인하도록 하는 등의 반복 작업에 귀중한 시간을 낭비하지 않도록 함으로써 네트워크 엔지니어링 리소스를 보다 생산적인 작업에 사용할 수 있습니다.

ZTP의 확장인 ZTR(Zero Touch Replacement)을 사용하는 경우에는 스위치를 물리적으로 교체할 수 있으며, 교체 스위치는 자동으로 교체 전의 원래 스위치와 동일한 이미지와 구성을 선택합니다. 스위치 ID 및 컨피규레이션은 스위치 MAC 주소가 아니라 장치가 연결된 네트워크 내의 위치에 연결됩니다(인접 장치의 LLDP 정보를 기준으로 함). 하드웨어 장애 및 RMA가 흔히 발생하지는 않겠지만, ZTR을 사용하는 경우 이러한 문제 상황을 해결하는 데 걸리는 복구 시간은 새 스위치가 도착해서 케이블을 물리적으로 연결하는 데 걸리는 시간으로 줄어듭니다. 또한 네트워크 엔지니어가 직렬 콘솔 케이블을 사용하여 스위치에 물리적으로 장치를 구성하지 않아도 됩니다.

VM TRACER

가상화된 데이터센터의 크기가 증가함에 따라 이러한 데이터센터를 지원하는 물리적/가상 네트워크의 크기 및 복잡성도 같이 증가했습니다. 가상 컴퓨터는 가상 스위치를 통해 연결한 다음 물리적 인프라에 연결되므로 추상화 레이어가 추가되고 네트워크가 더욱 복잡해집니다. 이와 관련하여 VMware 관리자는 새롭게 도입된 서버 쪽 도구를 통해 가상 컴퓨터와 네트워크를 관리할 수 있습니다. 그러나 물리적 네트워크와 가상 네트워크 간의 충돌을 해결할 수 있는 도구는 없었습니다.

Arista VM Tracer는 VMware vCenter API와 통신하여 가상화된 물리적 서버 그리고 물리적 서버가 포함되어야 할 VLAN(vCenter의 정책을 기준으로 사용)을 자동으로 검색한 다음 vMotion 이벤트를 통해 실시간으로 물리적 스위치 포트 구성을 자동 적용함으로써 이러한 문제를 해결합니다. 따라서 포트 구성 및 VLAN 데이터베이스 멤버 자격 확인을 자동으로 수행할 수 있으며, 트렁크 포트에서 VLAN을 동적으로 추가/제거할 수 있습니다.

또한 VM Tracer는 네트워크 기술자에게 물리적 스위치 포트의 물리적 서버와 VM을 자세히 파악하는 기능을 제공함과 동시에 서버 팀과 네트워크 팀이 작업을 유연하게 자동으로 수행할 수 있도록 합니다.

VXLAN

VXLAN은 대형 레이어 2 네트워크에서 기본적으로 발생하는 확장 문제 없이도 훨씬 규모가 큰 네트워크를 레이어 2에 구축할 수 있도록 하는 멀티 벤더 업계 지원 네트워크 가상화 기술입니다. VLAN과 비슷한 방식의 인캡슐레이션 기술을 사용하여 레이어 3에서 IP 패킷 내의 레이어 2 이더넷 프레임을 인캡슐레이션하며, 이것은 '오버레이' 네트워크로 분류됩니다. 가상 컴퓨터 측면에서 볼 때 VXLAN을 사용하면 물리적 서버가 있는 VLAN 또는 IP 서브넷에 관계없이 VM을 모든 위치의 어떤 서버에나 배치할 수 있습니다.

VXLAN은 레이어 2 네트워크 확장과 관련하여 기본적으로 발생하는 여러 가지 문제를 다음과 같이 해결합니다.

- 결합 도메인 증가 없이 대형 레이어 2 네트워크를 구축하도록 지원
- 4,000개의 VLAN 이상으로 확장 가능
- 여러 물리적 위치나 포트에 레이어 2 연결 가능
- 플러드(알 수 없는 목적지)를 지역화하고 트래픽을 단일 사이트로 브로드캐스트 가능
- 모든 장치가 다른 모든 MAC 주소를 확인할 필요 없이 대형 레이어 2 네트워크 구축 가능

VXLAN은 레이어 3에 걸친 레이어 2 오버레이를 지원하는 업계 표준 방법입니다. 여러 벤더가 VXLAN을 지원하므로 VXLAN을 배치하는 방법도 다양합니다. 예를 들어 VXLAN을 소프트웨어 기능으로 하이퍼바이저가 있는 가상 스위치에 배치하거나, 방화벽 및 부하 분산 어플라이언스에 배치하거나, L3 스위치에 내장된 VXLAN 하드웨어 게이트웨이에 배치할 수 있습니다. 하드웨어 VXLAN 게이트웨이 기능이 있는 Arista 스위치 플랫폼에는 'E' 또는 'X'자가 붙은 모든 Arista 스위치(Arista 7500E Series, Arista 7280E, Arista 7320X Series, Arista 7300E Series, Arista 7060X Series, Arista 7050X Series)와 Arista 7150 Series가 포함됩니다.

이러한 플랫폼은 Arista CloudVision, OVSDB와 같은 개방형 표준 기반의 비전매 프로토콜 또는 정적 구성을 통한 통합을 수행함으로써 유니캐스트 기반의 하드웨어 VXLAN 게이트웨이 기능을 지원합니다. 이처럼 하드웨어 VXLAN 게이트웨이 기능이 개방적으로 제공되므로 최종 사용자는 고객이 벤더에 고착화되지 않고도 적절한 클라우드 통합 플랫폼을 선택할 수 있습니다.

LANZ

Arista LANZ(레이턴시 분석기)를 사용하면 네트워크 혼잡으로 인해 성능 문제가 발생하기 전에 실시간으로 혼잡을 추적할 수 있습니다. 오늘날 사용되는 시스템은 사용자가 네트워크 속도 저하를 체감해야 혼잡을 감지하는 경우가 많습니다. 네트워크 팀이 문제 티켓을 수신해서 검사해야 중요한 인터페이스에서 패킷이 손실된 것을 확인할 수 있습니다. 이러한 상황에서 네트워크 팀이 기존에 선택할 수 있었던 가장 효율적인 해결 방법은 문제가 발생한 포트를 패킷 캡처 장치에 미러링하고 혼잡 문제가 다시 반복되기를 기다리는 방법이었습니다.

하지만 이제는 LANZ의 사전 혼잡 감지 및 알림 기능을 통해 관리자와 통합 어플리케이션이 모두 다음과 같은 작업을 수행할 수 있게 되었습니다.

- 레이턴시 또는 패킷 손실을 발생시키는 네트워크 상황을 미연에 방지
- 뚜렷하게 발생한 상황에 맞춰 어플리케이션 동작 조정
- 병목 현상 가능성을 조기에 격리하여 사전 용량 계획 수행
- 사후 처리 상관 관계 및 백 테스트용 과학적 데이터 유지

ARISTA EOS API

Arista eAPI(EOS API) 중 CAPI(명령 API)를 사용하면 안정적이며 쉽게 사용할 수 있는 구문을 통해 어플리케이션과 스크립트가 EOS를 프로그래밍 방식으로 완벽하게 제어할 수 있습니다. eAPI는 프로그래밍 방식 API를 통해 Arista 스위치의 모든 기능에 대한 모든 구성 명령 및 모든 상태를 표시합니다.

eAPI를 사용하도록 설정하면 스위치는 Arista CLI 구문을 사용하여 명령을 수락하며, JSON 형식으로 직렬화되고 HTTP/HTTPS를 통해 제공되는 컴퓨터 판독 가능 출력 및 오류로 응답합니다. 이처럼 프로토콜이 단순하고 모든 스크립팅 언어에 걸쳐 JSON-RPC 클라이언트를 사용할 수 있으므로 eAPI를 언어와 관계없이 활용할 수 있으며 기존 인프라와 업무 흐름에 쉽게 통합할 수 있습니다. 또한 제품에 기본적으로 포함된 스크립트나 기타 방식으로 제공되는 스크립트에서 eAPI를 사용할 수 있습니다.

명령의 구조화된 출력은 항상 향후 출시될 다수의 EOS 버전과 호환될 예정입니다. 그러므로 최종 사용자는 최신 EOS 버전으로 업그레이드하고 새로운 기능에 액세스하는 기능을 그대로 유지하면서 중차대한 어플리케이션을 안정적으로 개발할 수 있습니다.

OPENWORKLOAD

OpenWorkload는 물리적 환경과 가상 환경을 모두 완전하게 파악함으로써 개방형 방식으로 작업 부하를 이동하고, 뛰어난 가상화/통합 시스템과 통합하여 자동화를 진행하고, 문제 해결을 간소화해주는 네트워크 어플리케이션입니다.

- 원활한 확장 - 네트워크 가상화가 완전하게 지원되므로 주요 SDN 컨트롤러에 연결할 수 있습니다.
- 통합 - VMware NSX™, OpenStack, Microsoft, Chef, Puppet, Ansible 등과 연계되는 인터페이스가 제공되어 프로비저닝이 간단합니다.
- VM 레벨까지 작업 부하 파악 가능 - 정책을 이동하고 지속적으로 모니터링을 수행하고 클라우드 네트워크의 문제를 빠르게 해결할 수 있습니다.

VMware, OpenStack, Microsoft OMI와 통합이 가능하도록 설계된 Arista의 개방형 아키텍처는 모든 가상화 및 통합 시스템과 통합될 수 있습니다.

스마트 시스템 업그레이드

SSU(스마트 시스템 업그레이드) 기능을 사용하면 네트워크 업그레이드에 대한 부담이 줄어들어 어플리케이션 다운타임이 최소화되며 중차대한 변경 제어 시 발생하는 위험이 감소됩니다. SSU는 장치를 서비스에서 제외하거나 다시 포함할 수 있는 방식의 통합을 통해 Microsoft, F5, Palo Alto Networks 등의 데이터센터 인프라 협력사를 긴밀하게 연결하는 완전히 맞춤 구성 가능한 기능 세트를 제공합니다. 따라서 고객은 불필요한 다운타임 또는 시스템 정전 없이도 최신 소프트웨어 버전을 지속적으로 사용할 수 있습니다.

네트워크 원격 분석

결함 감지에서 결함 격리까지의 문제 해결 과정을 보다 빠르게 수행할 수 있는 새로운 모델인 네트워크 원격 분석은 언더레이 및 오버레이 네트워크 통계를 모두 포함하는 네트워크 상태 관련 데이터를 Splunk, ExtraHop, Corvil, Riverbed 등의 어플리케이션으로 스트리밍합니다. 그러면 중차대한 인프라 정보가 어플리케이션 레이어에 표시되므로 문제를 사전에 방지할 수 있습니다.

OPENFLOW 및 DIRECTFLOW

Arista EOS는 트래픽 필터링 및 리디렉션용으로 OpenFlow 컨트롤러를 통해 제어되는 OpenFlow 1.0을 지원합니다. 또한 Arista DirectFlow를 사용하여 SDN 어플리케이션(예: 탭 에그리제이션)으로 트래픽을 전달하는 controller-less 모드도 지원합니다. 따라서 프로덕션 네트워크에서 표준 IP 라우팅 프로토콜을 실행하면서 동시에 SDN 어플리케이션에 대해 특정 플로우 처리를 프로그래밍 방식으로 구성할 수 있습니다.

ARISTA EOS: 안정적이며 유동적인 플랫폼

Arista EOS(확장 가능 운영 체제)는 현존하는 최고급 네트워크 운영 체제라 할 수 있습니다. Arista EOS는 최신 소프트웨어 및 O/S 아키텍처, 투명한 방식으로 다시 시작 가능한 프로세스, 개방형 플랫폼 개발 기능, 수정되지 않은 Linux 커널, 그리고 스테이트풀 게시/사용을 위한 데이터베이스 모델 등 다양한 기능을 제공합니다.

EOS의 핵심 요소인 SysDB(시스템 데이터베이스)는 EOS의 모든 프로세스에 대해 상태를 저장하는 데 필요한 개체 모델을 기반으로 컴퓨터에서 생성되는 소프트웨어 코드입니다. EOS에서 수행되는 모든 프로세스 간 통신은 SysDB 개체에 대한 쓰기로 구현됩니다. 이러한 쓰기는 사용되는 에이전트로 전파되며, 해당 에이전트에서 이벤트가 트리거됩니다. 예를 들어 사용자 레벨 ASIC 드라이버는 포트에서 링크 장애를 감지하면 SysDB에 해당 내용을 씁니다. 그러면 LED 드라이버가 SysDB로부터 업데이트를 수신하여 포트 상태를 읽은 다음 LED 상태를 적절하게 조정합니다. 이처럼 시스템 전체로 상태를 전달하는 중앙식 데이터베이스 방식과 SysDB 코드가 자동으로 생성되는 방식은 위험과 오류를 줄입니다. 그러므로 소프트웨어 기능의 속도가 개선되며, 동일한 API를 사용할 수 있는 고객이 SysDB에서 알림을 받거나 스위치 기능을 맞춤 구성 및 확장할 수 있는 유연성이 제공됩니다.

Arista의 소프트웨어 엔지니어링 방법론을 사용하는 고객의 경우 품질과 일관성 측면에서 다음과 같은 이점도 얻을 수 있습니다.

- 사용자 공간에서 결함을 완벽하게 격리하고 SysDB를 통해 심각한 이벤트를 효율적으로 해결할 수 있습니다. 메모리 누수와 같이 흔히 발생하는 시나리오에서 시스템이 자가 복구됩니다. 모든 프로세스는 별도로 진행되며 IPC 또는 메모리 공유 없이 endian과는 독립적으로 진행되고 가능한 경우 멀티스레딩 방식을 사용합니다.
- 수동으로 소프트웨어를 테스트할 필요가 없습니다. 모든 테스트는 24시간 끊임없이 계속해서 자동 수행되며, Arista는 에뮬레이터 및 하드웨어에서 실행되는 운영 체제를 통해 프로토콜 및 단위 테스트 비용을 효율적으로 조정합니다.
- 모든 플랫폼에서 단일 시스템 바이너리를 유지할 수 있습니다. 따라서 각 플랫폼의 테스트 심도를 개선하고 출시 기간을 단축할 수 있으며 모든 플랫폼에서 기능 및 버그 해결 호환성을 유지할 수 있습니다.

EOS는 확장성이라는 핵심 개념을 실현할 수 있는 개발 플랫폼을 제공합니다. 이처럼 개방형 기반과 동급 최고의 소프트웨어 개발 모델을 통해 기능의 속도를 높이고, 업타임을 개선하고, 보수를 보다 쉽게 수행하고, 원하는 도구와 옵션을 선택할 수 있습니다.

ARISTA EOS 확장성

Arista EOS는 루트 레벨 관리자를 위한 완전한 Linux 셸 액세스 기능과 폭넓은 Linux 기반 도구 세트를 제공합니다. Arista가 추구하는 소프트웨어 원칙인 개방성에 따라 모든 SysDB 프로그래밍 모델 및 API 세트는 표준 Bash 셸을 통해 표시 및 제공됩니다. SysDB는 Arista에서 사용하는 API 중 제한적인 일부만 제공하는 폐쇄적인 API가 아닙니다. Arista 소프트웨어 개발자가 EOS 내의 주소 공간 사이에 사용하는 모든 프로그래밍 인터페이스는 제 3자 개발자와 Arista 고객에게도 제공됩니다.

Arista EOS 확장성을 맞춤 구성하고 사용하는 방법의 몇 가지 예는 다음과 같습니다.

- 모든 로그 파일을 매일 특정 NFS 또는 CIFS 공유에 백업하려는 경우 스위치에서 직접 스토리지를 장착한 다음 rsync 또는 rsnapshot 을 사용하여 구성 파일을 복사하면 됩니다.
- 인터페이스 통계 또는 LANZ 스트리밍 데이터를 스위치의 순차 순환 대기 방식의 데이터베이스에 저장하려는 경우 스위치에서 직접 MRTG 를 실행하면 됩니다.
- Internet2 PerfSonar 성능 관리 앱을 사용하려는 경우 로컬에서 바로 실행하면 됩니다.
- Nessus 를 실행하여 서버 부팅 시 보안 검사를 진행하려는 경우 올라오는 포트에 의해 트리거되는 이벤트 핸들러를 생성합니다.
- Chef, Puppet, CFEngine 또는 Sprinkle 을 사용하여 서버 환경을 자동화하려는 경우 이러한 프로그램 중 하나 또는 모든 프로그램을 사용하여 Arista 스위치 모니터링 및 구성도 자동화할 수 있습니다.
- 스위치에서 직접 서버를 PXE 부팅하려는 경우 스위치에서 바로 DHCP 및 TFTP 서버를 실행하면 됩니다.

EOS 가 작동하는 것과 같은 Linux 인스턴스에서 코드를 실행하기가 어려운 경우, 내장된 KVM 을 통해 스위치에서 게스트 "OS" 를 실행할 수 있습니다. CPU, RAM, vNIC 등의 리소스를 게스트 OS 에 할당할 수 있으며, Arista 는 엔터프라이즈급 SSD 를 통해 추가 플래시 스토리지가 포함된 스위치를 제공합니다.

기타 SDCN(소프트웨어 중심 클라우드 네트워킹) 기술

Arista SDCN(소프트웨어 중심 클라우드 네트워킹)에는 앞에서 대략적으로 설명한 EOS 기반 기술 외에도 자동화된 네트워크 설계를 확장할 수 있는 여러 가지 기술이 통합되어 있습니다. 이러한 기술의 몇 가지 예는 다음과 같습니다.

- AEM(고급 이벤트 모니터링)
- 자동화된 모니터링/관리
- Arista CloudVision

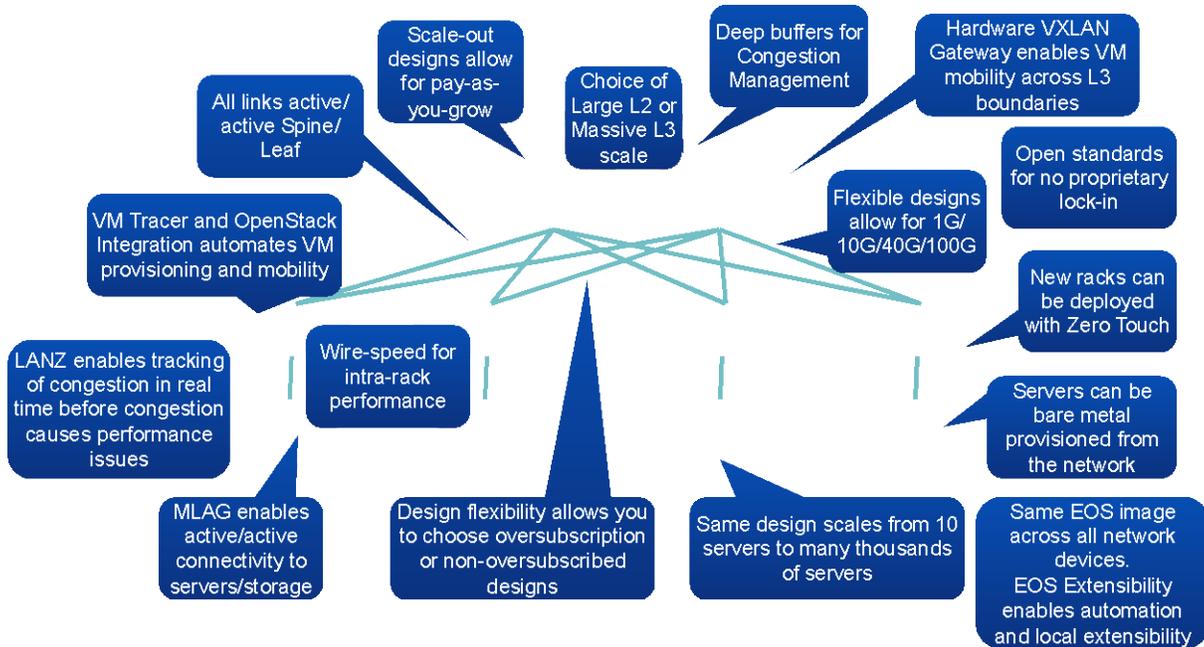


그림 21: Arista EOS 기반 기능 및 클라우드 네트워크 확장성

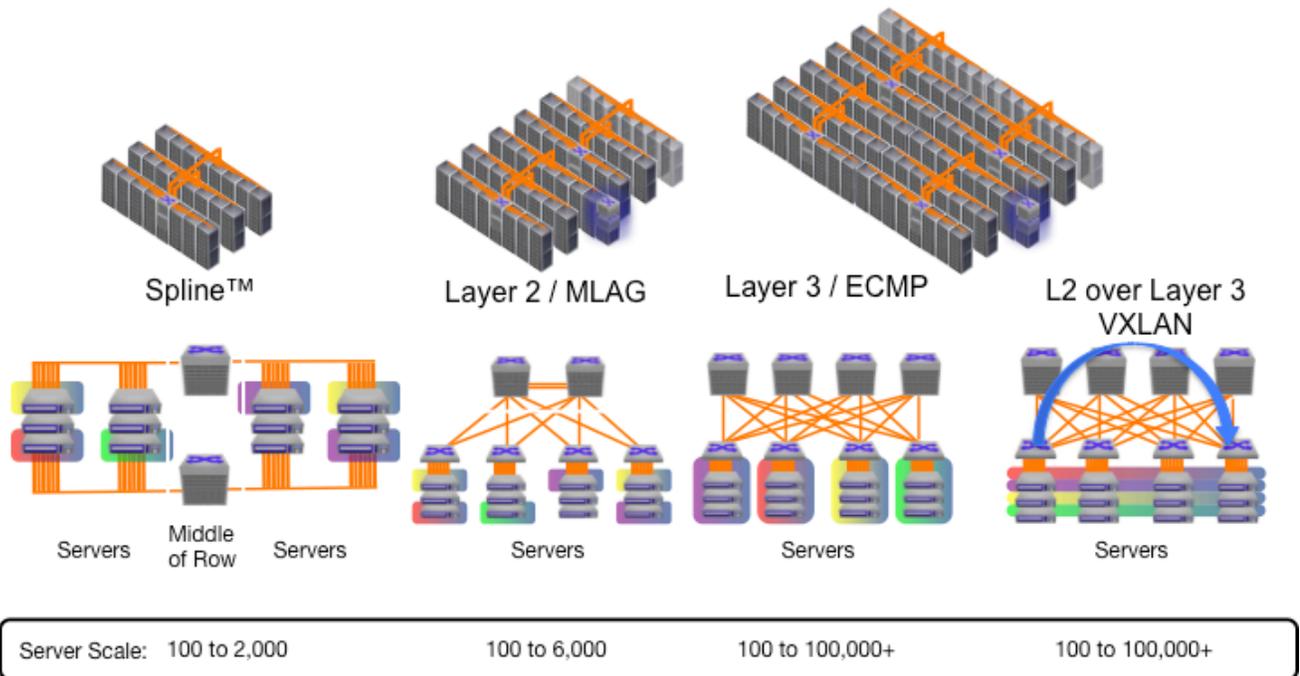


그림 22: Arista 클라우드 네트워크 설계: 단일 계층 Spline™ 및 2 계층 스파인/리프(포트 100 개~10 만 개 이상)

결론

Arista의 클라우드 네트워크 설계에는 정말 유용한 클라우드 컴퓨팅 기능(자동화, 자체 프로비저닝, 성능/비용이 일정하게 증가하는 확장 기능)을 제공한다는 원칙과 소프트웨어 정의 네트워킹의 원칙(네트워크 가상화, 맞춤형 프로그램 가능성, 간소화된 아키텍처, 보다 현실적인 가격)이 전매 또는 벤더 고착화 없는 방식이 채택되었습니다.

이러한 요소들을 통해 기업과 서비스 제공업체 데이터센터 모두에 네트워크의 가치를 최대화하는 동급 최고 소프트웨어의 기초를 확립합니다. 즉, 관리와 프로비저닝을 간소화하고, 서비스 제공 속도를 높이고, 비용을 줄이고, 경쟁력 있는 차별화 기회를 창출하는 동시에 네트워크 및 시스템 관리자에게 네트워크 제어 및 파악 능력을 다시 되돌려주는 새로운 아키텍처를 IT 인프라 내에서 가장 중차대한 위치에 구축할 수 있습니다.



산타 클라라 - 본사
5453 Great America Parkway
Santa Clara, CA 95054
전화: 408-547-5500
www.aristanetworks.com

샌프란시스코 - R&D 및 영업 사무소
1390 Market Street Suite 800
San Francisco, CA 94102

인도 - R&D 사무소
Eastland Citadel
102, 2nd Floor, Hosur Road
Madiwala Check Post
Bangalore - 560 095

밴쿠버 - R&D 사무소
Suite 350, 3605 Gilmore Way
Burnaby, British Columbia
Canada V5G 4X5

아일랜드 - 국제부 본사
Hartnett Enterprise Acceleration Centre
Moylish Park
Limerick, Ireland

싱가포르 - APAC 관리 사무소
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Arista Networks 정보

Arista Networks는 대형 데이터센터 및 컴퓨팅 환경을 위한 소프트웨어 중심 클라우드 네트워킹 솔루션을 제공하려는 목표로 설립되었습니다. 수상 경력에 빛나는 Arista 10 Gigabit Ethernet 스위치는 차원이 다른 확장성, 견고성, 가격 대비 성능을 제공합니다. 이러한 스위치를 통해 전 세계적으로 1 백만 개가 넘는 클라우드 네트워킹 포트가 배치되었습니다. Arista 플랫폼의 핵심 요소는 세계 최첨단의 네트워크 운영 체제인 EOS®(확장 가능 운영 체제)입니다. Arista Networks의 제품은 유통 협력사, 시스템 통합업체, 재판매인을 통해 전 세계적으로 제공됩니다.

추가 정보와 리소스는 www.aristanetworks.com 에서 확인할 수 있습니다.