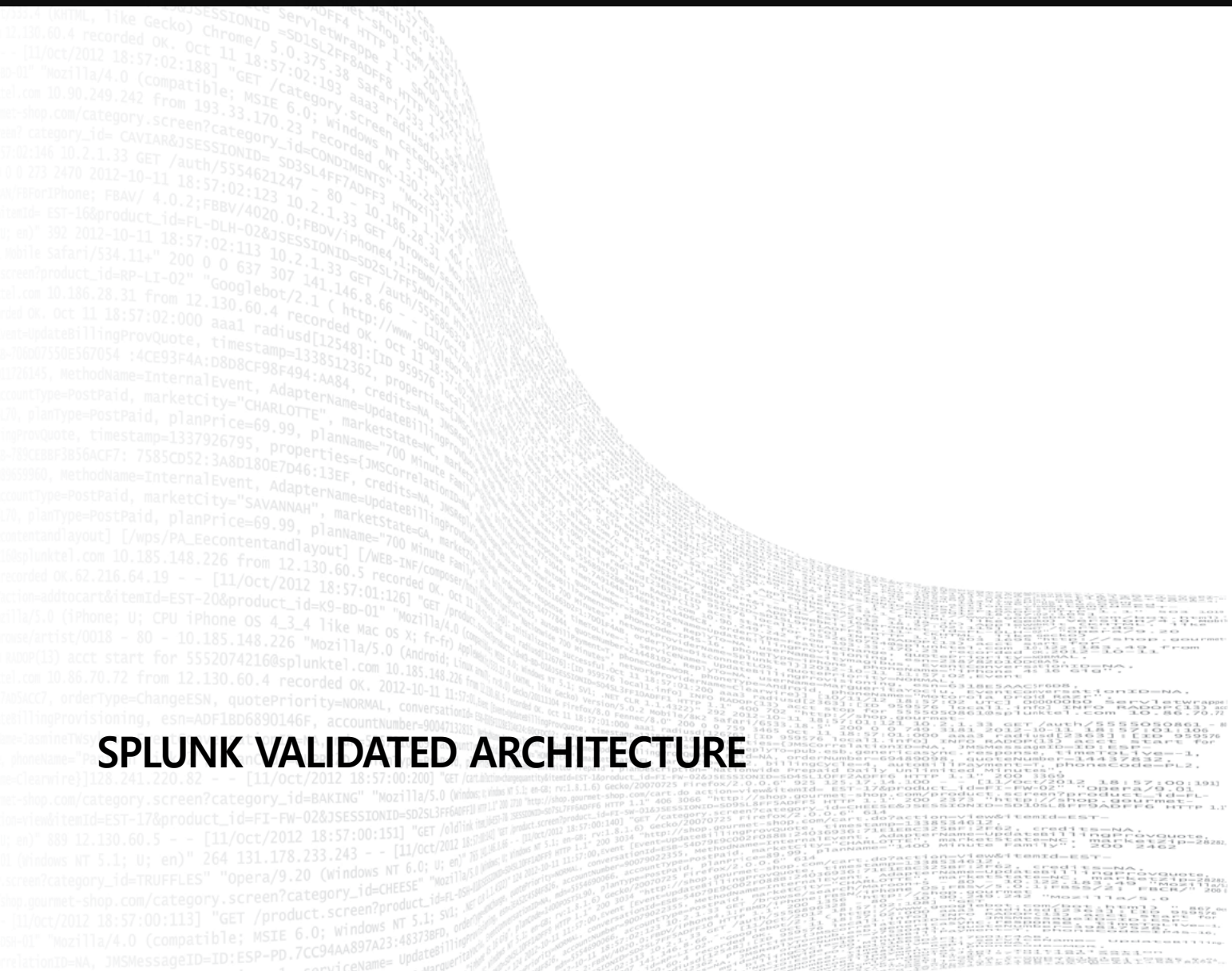


SPLUNK VALIDATED ARCHITECTURE

목차

소개 2
문서 구조 2
Splunk Validated Architecture를 사용하는 이유 3
Splunk Validated Architecture의 기본 요소 3
Splunk Validated Architecture의 기대 효과 4
역할 및 책임 4
Splunk Validated Architecture 선별 프로세스 개요 5
1a단계: 인덱싱 및 검색 요구 사항 정의 6
2a단계: 인덱싱 및 검색 토폴로지 선택 12
1b단계: 데이터 컬렉션 요구 사항 정의 24
2b단계: 데이터 컬렉션 구성 요소 선택 29
3단계: 설계 원리 및 모범 사례 적용 44
요약 및 다음 단계 53
다음 단계 53
부록 54
부록 "A": SVA 기본 요소 설명 54
부록 "B": 토폴로지 구성 요소 55

소개

Splunk Validated Architecture(SVA)는 안정적이고 효율적이며 반복 가능한 Splunk 배포를 위한 검증된 참조 아키텍처입니다. Splunk의 기존 고객 중 상당수는 빠른 도입과 확장을 경험함에 따라 확장을 시도해야 하는 과제에 직면했습니다. 동시에 초기 배포가 견고한 기반에서 구축될 수 있도록 지침과 인증된 아키텍처를 찾는 Splunk의 신규 고객이 점차 늘어나고 있습니다. SVA는 이렇게 증가하는 고객의 요구를 충족하기 위해 개발되었습니다.

SVA는 Splunk의 기존 고객이나 신규 고객에 관계없이 모든 사용자가 더 쉽게 유지 관리하고 간단하게 문제를 해결할 수 있는 환경을 구축할 수 있도록 지원합니다. SVA는 총소유비용을 최소화하면서 최적의 결과를 얻을 수 있도록 고안되었습니다. 또한 전체 Splunk 기반은 반복 가능한 아키텍처를 기반으로 하므로 시간 경과에 따라 필요한 경우 배포를 확장할 수 있습니다.

SVA는 다양한 조직의 요구 사항을 고려한 토폴로지 옵션을 제공하므로 사용자는 요구 사항에 적합한 토폴로지를 쉽게 이해하고 찾을 수 있습니다. Splunk Validated Architecture 선별 프로세스는 특정 요구 사항을 조직의 요구 사항에 가장 적합한 토폴로지와 일치시키는 데 도움이 됩니다. Splunk를 처음 사용하는 경우 초기 배포를 위해 Validated Architecture를 구현하는 것이 좋습니다. 기존 고객인 경우 Validated Architecture 토폴로지와 정렬하는 옵션을 탐색하는 것이 좋습니다. 사용자 지정 아키텍처를 구축해야 하는 고유한 요구 사항이 없다면 Validated Architecture로 비용 효율성을 유지하면서 요구 사항을 쉽게 충족할 수 있습니다.

이 백서는 SVA에 대한 개요를 제공합니다. 이 백서에는 요구 사항 설문지, 배포 토폴로지 다이어그램, 설계 원리 및 일반 지침을 비롯하여 SVA 선별 프로세스를 진행하는 데 필요한 리소스가 나와 있습니다.

Splunk Validated Architecture를 구현하는 데 도움이 필요한 경우 Splunk 프로페셔널 서비스(https://www.splunk.com/en_us/support-and-services/splunk-services.html)로 문의하십시오.

문서 구조

SVA는 세 가지 주요 콘텐츠 분야로 나뉩니다.

1. 인덱싱 및 검색 토폴로지
2. 데이터 컬렉션 아키텍처 구성 요소
3. 설계 원리 및 모범 사례

인덱싱 및 검색은 Splunk 배포의 핵심 인덱싱 및 검색 기능을 제공하는 아키텍처 계층을 포괄합니다. 데이터 컬렉션 구성 요소 절에서는 요구 사항에 맞는 올바른 데이터 컬렉션 메커니즘을 선택하는 데 필요한 지침을 제공합니다.

설계 원리 및 모범 사례는 아키텍처 전반에 적용되며 배포 세부 작업을 수행할 때 올바른 선택을 할 수 있도록 지원하는 데 유용합니다.

Splunk Validated Architecture를 사용하는 이유

Validated Architecture를 구현하면 Splunk를 한층 더 확실하게 설계 및 배포할 수 있습니다. SVA는 다음과 같은 조직이 직면한 가장 공통적인 과제 중 일부를 해결하는 데 유용합니다.

성능

- 조직은 성능과 안정성을 향상하길 원합니다.

복잡성

- 조직이 급속도로 성장하거나 유기적으로 성장한 경우 사용자 지정 구축 배포라는 위험을 감수하기도 합니다. 그러한 경우 배포 환경이 불필요하게 복잡해질 수 있습니다. 이러한 복잡성은 확장을 시도할 때 심각한 방해 요소가 될 수 있습니다.

효율성

- 조직은 Splunk 배포의 이점을 최대로 활용하기 위해 운영 효율성을 향상시키고 더 빠르게 가치를 창출해야 합니다.

비용

- 조직은 TCO(총소유비용)를 절감하는 동시에 모든 요구 사항을 충족할 수 있는 방법을 모색하고 있습니다

민첩성

- 조직은 확장과 성장에 따른 변화에 적응해야 합니다.

메인터넌스

- 메인터넌스에 드는 수고를 줄이기 위해 환경을 최적화해야 합니다

확장성

- 조직은 효율적이고 원활하게 확장할 수 있어야 합니다.

확인

- 조직 내의 이해 관계자는 Splunk 배포가 모범 사례를 기반으로 구현되길 원합니다.

Splunk Validated Architecture의 기본 요소

Splunk Validated Architecture는 다음과 같은 기본적인 요소를 기반으로 설계되었습니다. 이러한 설계 기본 요소에 대한 자세한 내용은 아래의 "부록 A"를 참조하십시오.

가용성	**성능**	**확장성**	**보안**	**관리 기능**
시스템은 **지속적으로 작동 가능하며** 계획 및 계획되지 않은 작동 중단으로부터 복구할 수 있습니다.	시스템은 다양한 사용 패턴에서 **최적의 서비스 수준을 유지**할 수 있습니다.	시스템은 모든 계층에서 확장할 수 있도록 설계되어 **증가된 워크로드를 효율적으로 처리**할 수 있습니다.	시스템은 **데이터, 설정 및 자산을 보호하도록 설계되었으며** 지속적으로 가치를 제공합니다.	시스템은 **모든 계층에서 중앙 집중식으로 작동 및 관리할 수 있습니다.**

이 기본 요소는 Splunk Center Of Excellence 모델의 **플랫폼 관리 및 지원** 서비스에 대한 직접 지원을 제공합니다.

Splunk Validated Architecture의 기대 효과

SVA에는 배포 기술이나 배포 규모 조정이 포함되어 있지 않습니다. 이에 대한 근거는 다음과 같습니다.

- 운영 체제 및 서버 하드웨어 같은 배포 기술은 SVA 환경에서 선택한 구현으로 간주됩니다. 고객마다 다양한 선택 옵션이 있으므로 일반화하는 것은 쉽지 않습니다.
- 배포 규모 조정을 위해서는 데이터 수집 볼륨, 데이터 유형, 검색 볼륨 및 검색 이용 사례에 대한 평가가 필요하며, 이 사례는 매우 고객 중심적이며 일반적으로 기본적인 배포 아키텍처 자체와 관련이 없는 경향이 있습니다. 기존의 규모 조정 도구는 배포 아키텍처를 설정하여 이 프로세스를 지원합니다. Splunk 스토리지 규모 조정(https://splunk-sizing.appspot.com/)은 사용 가능한 도구 중 하나입니다.

SVA가 제공하는 항목:	**SVA가 제공하지 않는 항목:**
✅ 클러스터된 배포 옵션 및 클러스터되지 않은 배포 옵션 ✅ 참조 아키텍처 다이어그램 ✅ 사용자에게 적합한 아키텍처를 선택할 수 있도록 도움을 주는 지침 ✅ 계층별 권장 사항 ✅ Splunk 배포를 개발하기 위한 모범 사례	❌ 구현 선택 사항(OS, 베어메탈, 가상, Cloud 등 비교). ❌ 배포 규모 조정 ❌ 아키텍처에 대한 규범적 승인 참고: SVA는 궁극적으로 조직에 가장 적합한 결정을 내릴 수 있도록 권장 사항과 지침을 제공합니다. ❌ 가능한 모든 배포 시나리오에 대한 토폴로지 제안. 경우에 따라 고유한 요소로 인해 사용자 지정 아키텍처를 개발해야 할 수도 있습니다. Splunk 전문가는 필요한 사용자 지정 솔루션을 제공할 수 있습니다. 기존 고객인 경우 Splunk 계정 팀에 문의하십시오. Splunk를 처음 사용할 경우 여기(https://www.splunk.com/en_us/talk-to-sales.html)로 문의하실 수 있습니다.

역할 및 책임

Splunk Validated Architecture는 의사 결정권자 및 관리자의 우려 사항과 깊은 관련이 있습니다. 엔터프라이즈 아키텍트, 컨설턴트, Splunk 관리자 및 관리 서비스 공급업체는 모두 SVA 선별 프로세스에 참여해야 합니다. 아래에서 각 역할에 대한 설명을 확인할 수 있습니다.

역할	설명
엔터프라이즈 아키텍트	엔터프라이즈 요구를 충족시키기 위해 Splunk 배포 아키텍트를 담당합니다.
컨설턴트	Splunk 아키텍처, 설계 및 구현을 위한 서비스를 제공합니다.
Splunk 엔지니어	Splunk 라이프사이클을 관리합니다.
매니지드 서비스 공급자	고객용 서비스로 Splunk를 배포 및 실행하는 엔티티를 담당합니다.

Splunk Validated Architecture 선별 프로세스 개요

Splunk Validated Architecture 선별 프로세스는 조직에 대한 모든 요구 사항을 충족하는 가장 단순하고 간소화된 아키텍처를 식별하는 데 도움이 됩니다.

선별 프로세스 단계	목표	고려 사항
1단계: 요구 사항 정의: a) 인덱싱 및 검색 b) 데이터 컬렉션 메커니즘	요구 사항 정의	• 의사 결정권자, 이해 관계자 및 관리자는 조직의 요구 사항을 식별하고 정의하기 위해 협력해야 합니다. • 이미 배포한 경우 현재 아키텍처를 평가하여 검증된 모델로 전환하는 데 필요한 조치를 확인할 수 있습니다. 요구 사항을 정의하는 데 도움이 되는 설문지는 아래 1단계를 참조하십시오.
2단계: 토폴로지 선택: a) 인덱싱 및 검색 b) 각 데이터 컬렉션 메커니즘	확인된 요구 사항을 충족하는 토폴로지를 선택하십시오.	• 요구 사항을 충족하는 가장 적합한 토폴로지를 선택합니다. • SVA에 따라 간단하게 설명하면 한층 더 쉽게 확장성에 대해 이해할 수 있습니다.

		토폴로지 옵션에 대한 다이어그램 및 설명은 아래의 2단계를 참조하십시오.
3단계: 설계 원리 및 모범 사례 적용	설계 원리의 우선 순위를 지정하고 계층별 구현 모범 사례를 검토합니다.	• 각 설계 원리는 Splunk Validated Architecture의 하나 이상의 기본 요소를 보강합니다. • 조직의 요구에 따라 설계 원리의 우선 순위를 결정합니다. • 계층별 권장 사항은 토폴로지 구현 시 필요한 지침을 제공합니다. 설계 원리에 대한 세부 사항은 아래 3단계를 참조하십시오.

1a단계: 인덱싱 및 검색 요구 사항 정의

적절한 배치 토폴로지를 선택하려면 요구 사항을 상세히 검토해야 합니다. 요구 사항을 정의한 후 Splunk를 배포하기 위한 가장 단순하고 비용 효율적인 방법을 선택할 수 있습니다. 아래에는 배포에 대한 인덱싱 및 검색 계층에 대한 주요 요구 사항 영역을 정의하는 데 유용한 설문지가 나와 있습니다.

요구 사항 설문지는 배포 토폴로지에 직접적인 영향을 미치는 영역에 중점을 둡니다. 따라서 다음 단계에서 토폴로지를 선택하기 전에 아래 질문에 대한 대답을 기록하는 것이 좋습니다.

고려 사항

이용 사례 검토

요구 사항을 정의할 때 Splunk 인프라의 사용 목적을 고려해야 합니다. 예를 들어 부서별 DevOps 이용 사례에 대한 토폴로지는 종종 몇몇 경우에 미션 크리티컬 이용 사례보다 간단합니다. 다음과 관련된 이용 사례를 충분히 고려해야 합니다.

- 검색
- 가용성
- 규정 준수 요구 사항(항상 100% 데이터 충실도 및 가용성을 유지해야 하는 경우 특히 중요함)
- 조직과 관련된 기타 이용 사례 시나리오

이용 사례 시나리오에 따라 배포 시 추가 아키텍처 특성을 제공해야 할 수 있습니다.

미래 성장에 대한 고려

요구 사항을 정의하려면 즉각적인 요구 사항을 고려해야 합니다. 그러나 향후 성장 및 확장성도 고려해야 합니다. 배포 규모를 확장하려면 현재 계획할 지출, 추가 인력 또는 기타 리소스가 필요할 수 있습니다.

토폴로지 범주

다음은 SVA 토폴로지 범주의 핵심입니다. 이 범주는 아래의 설문지에서 사용됩니다. 또한 SVA 선별 프로세스의 다음 단계에서 이러한 범주에 대한 참조를 찾을 수 있습니다.

인덱싱 계층 범주

범주 코드	설명
S	범주 "S"는 단일 서버 Splunk 배포의 인덱서를 나타냅니다.
D	범주 "D"는 최소 2개의 인덱서가 있는 분산형 인덱서 계층의 필요성을 나타냅니다.
C	범주 "C"는 클러스터된 인덱서 계층의 필요성을 나타냅니다(데이터 복제가 필요함).
M	범주 "M"은 여러 사이트를 포함하는 클러스터된 인덱서 계층의 필요성을 나타냅니다.

검색 계층 범주

범주 코드	설명
1	범주 "1"은 단일 검색 헤드가 요구 사항을 충족할 수 있음을 나타냅니다.
2	범주 "2"는 여러 검색 헤드가 요구 사항을 충족해야 함을 나타냅니다.
3	범주 "3"은 검색 헤드 클러스터가 요구 사항을 충족해야 함을 나타냅니다.
4	범주 "4"는 요구 사항을 충족시키기 위해 여러 사이트("확장된" SHC)에 있는 검색 헤드 클러스터를 나타냅니다.
+10	범주 "+10"은 Enterprise Security App을 지원하기 위해 전용 검색 헤드(클러스터)가 필요함을 나타냅니다. 검색 계층 토폴로지 범주에 10을 추가하고 이 앱의 특정 요구 사항에 대한 토폴로지 관련 설명을 주의 깊게 읽으십시오.

설문지 1: 인덱스 및 검색 계층 요구 사항 정의

♦ 토폴로지 범주 코드에 대한 설명은 위의 키를 참조하십시오. 여러 질문에 "예"라고 답변한 경우 가장 높은 번호가 지정된 질문에 토폴로지 범주 코드를 사용합니다.

#	질문	고려 사항	토폴로지에 미치는 영향	인덱서 계층 토폴로지 범주 ♦	검색 계층 토폴로지 범주 ♦
1	예상 일일 데이터 수집량이 300GB/일 미만입니까?	일일 수집량의 단기적 성장 고려(~6-12개월)	가용성 관련 질문에 대한 답변에 따라 단일 서버 배포를 위한 후보	S	1
2	데이터 컬렉션/인덱스 처리를 위해 고가용성이 필요합니까?	지속적인 데이터 수집이 필요한 이용 사례를 모니터링하기 위해 Splunk를 사용할 계획이 없다면 인바운드 데이터 흐름이 일시적으로 중단될 수 있습니다. 이때 로그 데이터가 손실되지 않는다고 가정합니다.	지속적인 수집을 지원하는 분산 배포 필요	D	1
3	사용 가능한 검색 헤드로 검색을 실행한다고 가정합니다. 데이터가 항상 검색 가능해야 합니까? 예를 들어, 검색 결과의 완성도에 영향을 줄 수 없습니까?	예를 들어 이용 사례에서 집계 함수를 사용하여 성능 메트릭 및 일반 사용량 모니터링을 계산하는 경우 단일 인덱서의 작동 중단은 많은 수의 이벤트에 대한 통계 계산에 실제로 영향을 미치지 않을 수도 있습니다. 이용 사례가 보안 감사 및 위협 탐지에	복제 팩터가 둘(2) 이상인 클러스터된 인덱서가 필요합니다. 참고: 복제 팩터 2는 단일 인덱서 노드 오류에 대한 보호 수준이 가장 낮지만 권장 (및 기본값) 복제 팩터는 3입니다.	C	1

#	질문	고려 사항	토폴로지에 미치는 영향	인덱서 계층 토폴로지 범주 ♦	검색 계층 토폴로지 범주 ♦
		대한 내용을 다루는 경우 검색 결과에 사각 지대 같은 영역이 발생하는 것은 바람직하지 않을 수 있습니다.			
4	데이터 센터 작동 중단 시 여러 데이터 센터를 운영하고 Splunk 환경을 자동으로 복구해야 합니까?	재해 복구 요구 사항은 수동 재해 복구를 위해 두 가지 시설(활성/활성)에서 지속적으로 운영하거나 RTO/RPO 목표를 수행하도록 지시할 수 있습니다.	지속적으로 운영하려면 데이터 수집/인덱싱 계층과 검색 계층에서 페일오버를 보장하는 멀티 사이트 인덱서 클러스터링과 두 개 이상의 활성 검색 헤드가 필요합니다.	M	2
5	연속적인 무손실 데이터 수집을 가정할 경우 사용자 중심 검색 계층에 HA가 필요합니까?	Splunk를 지속적인 단기 모니터링에 사용할 경우 검색 계층의 작동 중단은 허용될 수 없습니다. 이는 다른 이용 사례에 해당되지 않을 수도 있습니다.	중복 검색 헤드와 잠재적으로 검색 헤드 클러스터링이 필요합니다.	D/C/M	3
6	다수의 동시 사용자 및/또는 중요한 예약된 검색 워크로드를 지원해야 합니까?	51명 이상의 동시 사용자/검색에 대한 요구 사항은 일반적으로 검색 계층의 수평 확장이 필요합니다.	검색 계층에서 검색 헤드 클러스터를 사용하는 토폴로지가 필요할 수 있습니다.	D/C/M	3
7	여러 데이터 센터 환경에서 사이트 간	사이트 작동 중단 시 사용자가 현재와	적절한 설정이 지정된 전체	M	4

#	질문	고려 사항	토폴로지에 미치는 영향	인덱서 계층 토폴로지 범주 ♦	검색 계층 토폴로지 범주 ♦
	사용자 아티팩트(검색, 대시보드 및 기타 knowledge object)를 동기화해야 합니까?	일관된 환경을 유지할지 여부를 결정합니다.	사이트에 걸쳐 "확장된" 검색 헤드 클러스터가 필요합니다. **중요 사항:** 확장된 SHC는 전체 사이트에서 오류가 발생하는 동안 사용자의 검색 가용성을 향상시킬 수 있지만 일부 아티팩트는 두 사이트에서 복제되지 않을 수 있습니다. 이 경우 Splunk App for Enterprise Security 같이 현재 일관된 아티팩트에 의존하는 특정 애플리케이션에 영향을 미칠 수 있습니다. 검색 헤드 클러스터링만 사용해서는 완벽한 DR 솔루션을 제공할 수 없습니다. SHC의 다른 이점도 여전히 적용됩니다.		
8	Splunk App for Enterprise Security(ES)를 배포하시겠습니까?	각 토폴로지와 함께 기술된 바와 같이 Splunk App for Enterprise Security에 특정 제한 사항이	ES에는 전용 검색 헤드 환경(독립형 또는 클러스터형)이 필요합니다.	D/C/M	+10

#	질문	고려 사항	토폴로지에 미치는 영향	인덱서 계층 토폴로지 범주 ♦	검색 계층 토폴로지 범주 ♦
		적용됨을 읽고 숙지하십시오.			
9	데이터 관리 규정이 적용되는 지리적으로 분산된 환경이 있습니까?	일부 국가에서는 규정에 따라 해당 국가에서 생성된 데이터가 각 국가의 시스템을 벗어나는 것을 허용하지 않습니다.	이러한 규정은 중앙식 Splunk 인덱싱 계층 배포를 방지하여 심층적인 배포와 같은 세부 사항을 고려하는 Splunk/파트너사, 고객 간 협업을 통해 사용자 지정 아키텍처를 개발하도록 요구합니다. 즉, 이러한 요구 사항을 충족시키는 SVA는 없습니다.	사용자 지정	사용자 지정
10	공유 서버/인덱서에서 특정 로그 데이터 원본의 코로케이션(co-location)을 방지하는 매우 제한적인 보안 정책이 있습니까?	중요도가 높은 로그 데이터는 회사 정책에 따라 동일한 물리적 시스템/동일한 네트워크 영역 내에서 위험도가 낮은 데이터 집합과 함께 보관할 수 없습니다.	잠재적으로 공유 하이브리드 검색 계층을 사용하여 여러 개의 독립적 인덱싱 환경이 필요합니다. 이 경우 SVA의 범위를 벗어나며 사용자 지정 아키텍처 개발이 필요합니다.	사용자 지정	사용자 지정

토폴로지 범주 코드를 확인하는 방법

위의 요구 사항 설문지의 답변에 따라 필요성에 대한 최적의 토폴로지를 식별하도록 지원하는 토폴로지 범주 지표가 통합됩니다. 관련 지침과 예시는 다음과 같습니다.

지침

1. "예"라고 답변한 질문을 적습니다.
2. 여러 질문에 "예"라고 답변한 경우 가장 높은 번호가 지정된 질문에 대해 토폴로지 권장 사항을 따릅니다. 여러 토폴로지 옵션(예: "D/C/M")이 표시되면 이전 질문을 확인하고 어떤 옵션이 가장 적합한지 살펴봅니다.
3. 토폴로지 범주 코드는 인덱서 계층을 나타내는 문자로 시작합니다(예: "C" 또는 "M"). 이 문자 다음에는 검색 계층을 나타내는 숫자(예: "1" 또는 "13")가 표시됩니다.

예시 1번:

3번, 5번 및 8번 질문에 "예"라고 답변했다고 가정합니다. 이 경우 토폴로지 범주는 "C13"이 되며, 이는 두 개의 검색 헤드 클러스터가 있는 클러스터된 인덱싱 계층이 필요하다는 것을 나타냅니다.

예 2:

1번 질문에만 "예"라고 답변했다고 가정합니다. 토폴로지 범주는 "S1"이 되며, 이는 단일 서버 Splunk 배포가 이상적인 토폴로지라는 것을 나타냅니다.

2a단계: 인덱싱 및 검색 토폴로지 선택

토폴로지는 일반적으로 클러스터되지 않은 배포와 클러스터된 배포로 나뉩니다. 클러스터되지 않은 배포의 경우 최소한의 고유 구성 요소가 필요하며 우수한 확장성을 가집니다. 클러스터되지 않은 배포의 경우 가용성 및 재해 복구 기능이 부족해지는 반면, 배포 옵션은 여전히 조직에 적합한 선택일 수 있습니다.

주의사항: SVA 선별 프로세스의 주요 목표는 불필요한 구성 요소를 배제하고 필요한 요소를 채택하여 구축하는 것입니다.

참고

즉각적인 요구 사항 외에도 추가적인 이점을 제공하는 토폴로지를 구현하도록 선택할 수 있지만 불필요한 비용이 발생할 수 있습니다. 게다가 추가적인 복잡성이 도입되면 운영 효율성을 낮추는 비생산적인 경우가 있습니다.

토폴로지 다이어그램에 대한 중요 참고 사항

토폴로지 다이어그램의 아이콘은 **기능별 Splunk 역할**을 나타내며, 이를 실행하기 위한 전용 인프라를 의미하지 않습니다. Splunk 역할을 동일한 인프라/서버에 코로케이션할 수 있는 지침을 보려면 부록을 참조하십시오

토폴로지 범주 코드 사용

토폴로지 옵션을 선택하기 전에 요구 사항 설문지를 작성하여 토폴로지 범주 코드를 결정하는 것이 가장 좋습니다. 이 작업을 수행하지 않은 경우 뒤로 이동하여 위의 이전 단계를 완료하십시오. 토폴로지 범주 코드가 있으면 명시된 요구 사항에 가장 적합한 배포 옵션을 파악할 수 있습니다.

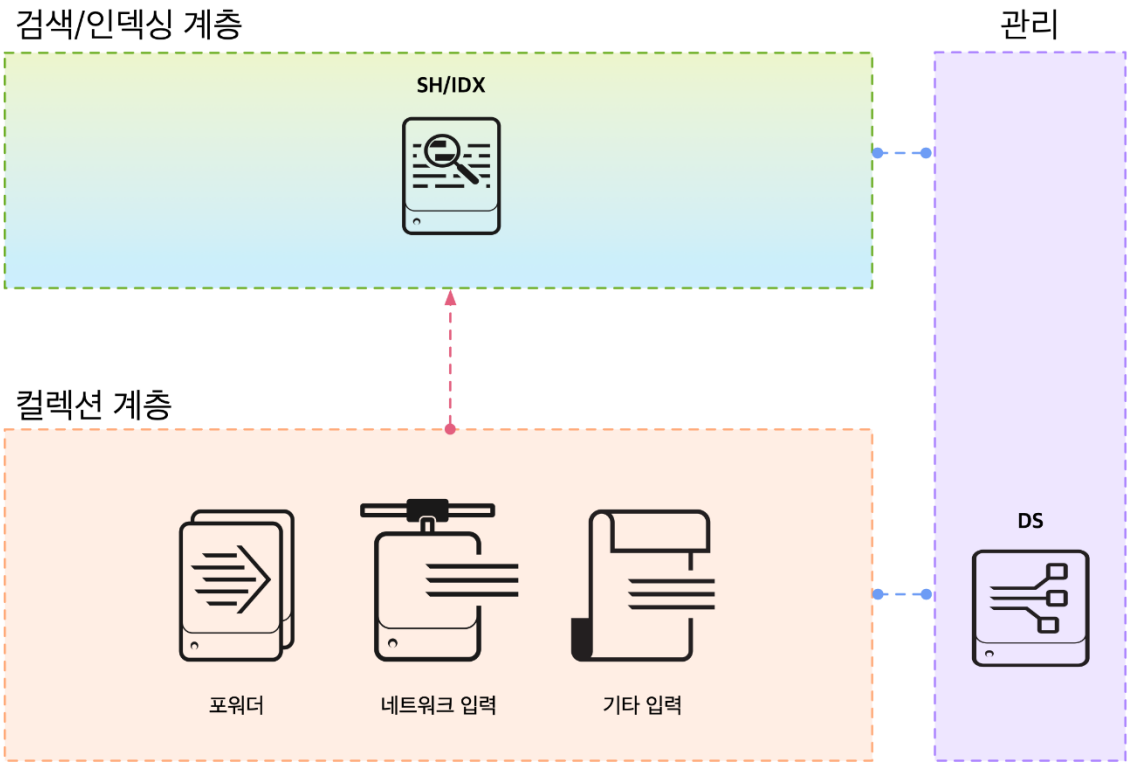
클러스터되지 않은 배포 옵션

아래에서 다음과 같은 토폴로지 옵션을 찾을 수 있습니다.

배포 유형	토폴로지 범주 코드
단일 서버 배포	S1
클러스터되지 않은 분산 배포	D1/D11

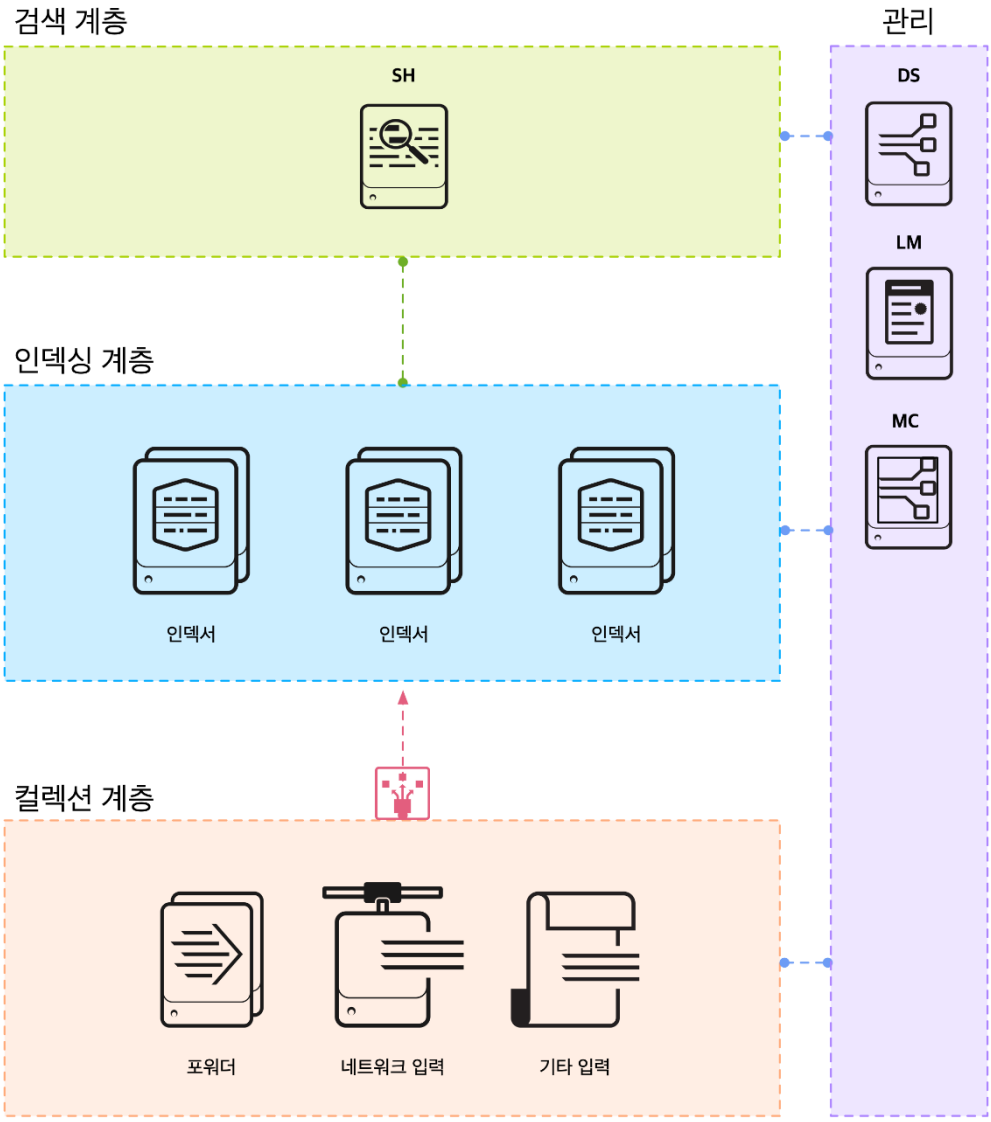
토폴로지 구성 요소에 대한 설명은 아래의 부록 "B"를 참조하십시오.

단일 서버 배포(S1)



단일 서버 배포 설명(S1)	제한 사항
이 배포 토폴로지는 사용자 환경이 다음 기준을 모두 충족하는 경우 매우 경제적인 솔루션을 제공합니다. a) Splunk 배포에 고가용성 또는 자동 재해 복구 기능을 제공할 필요가 없음, b) 일일 데이터 수집량이 300GB/일 미만, c) 중요하지 않은 검색 이용 사례를 가진 소수의 사용자를 보유함. 이 토폴로지는 일반적으로 비즈니스 크리티컬하지 않은 소규모의 이용 사례(일반적으로 부서별)에 사용됩니다. 적절한 이용 사례에는 데이터 온보딩 테스트 환경, 소형 DevOps 이용 사례, 애플리케이션 테스트와 통합 환경 및 이와 유사한 시나리오가 포함됩니다. 이 토폴로지의 주요 이점으로는 간편한 관리, 소량 데이터 볼륨에 대한 우수한 검색 성능 및 고정 TCO가 있습니다.	• 검색/인덱싱을 위한 고가용성 없음 • 하드웨어 용량에 따라 제한되는 확장성(분산 배포로의 직접적인 마이그레이션 경로)

클러스터되지 않은 분산 배포(D1/D11)



클러스터되지 않은 분산 배포에 대한 설명(D1/D11)	제한 사항
다음 상황 중 하나의 분산 토폴로지로 이동해야 합니다. a) Splunk에 전송할 일일 데이터 볼륨이 단일 서버 배포 용량을 초과하거나 b) 고가용성의 데이터 수집을 원하거나 필요로 하는 경우. 여러 개의 독립 인덱서를 배포하면 인덱싱 용량을 선형으로 확장하고 데이터 수집의 가용성을 암시적으로 높일 수 있습니다. 인덱서 노드를 추가하면 TCO가 예측 가능한 선형 방식으로 증가합니다. 권장하는 Monitoring Console(MC) 구성 요소를 도입하면 분산 배포의 상태 및 용량을 모니터링할 수 있습니다 또한 MC는 중앙 집중식 경고 시스템을 제공하므로 배포에서 비정상적인 상태에 대한 알림을 수신합니다.	• 검색 계층을 위한 고가용성 없음 • 인덱싱 계층의 제한된 고가용성, 노드 오류로 인해 과거 검색에 대한 검색 결과가 불완전하게 나타날 수 있습니다.

클러스터되지 않은 분산 배포에 대한 설명(D1/D11)	제한 사항
새로운 인덱서를 추가할 때마다 검색 헤드를 사용 가능한 검색 피어 리스트와 함께 수동으로 설정해야 합니다. **ES 고객 참고 사항:** 범주 코드가 D1(예: Splunk App for Enterprise Security를 배포하려는 경우)인 경우 **전용 단일** 검색 헤드는 앱을 배포하는 데 필요합니다(토폴로지 다이어그램에는 표시되지 않음). 컬렉션 계층은 새 인덱서가 추가될 때마다 (배포 서버를 통해) 대상 인덱서 리스트로 설정해야 합니다. 이 배포 토폴로지는 1000개가 넘는 인덱서 노드로 선형 확장될 수 있으므로 매우 많은 양의 데이터 수집 및 검색 볼륨을 지원할 수 있습니다. 여러 인덱서에서 병렬 검색을 실행하여 전체 대규모 데이터 집합에서 검색 성능을 유지할 수 있습니다(매핑/축소). 별도의 토폴로지로 분류되지 않지만 검색 헤드 클러스터를 사용하여 검색 계층에서 검색 용량을 늘릴 수 있습니다(토폴로지 C3/C13의 검색 계층 참조).	

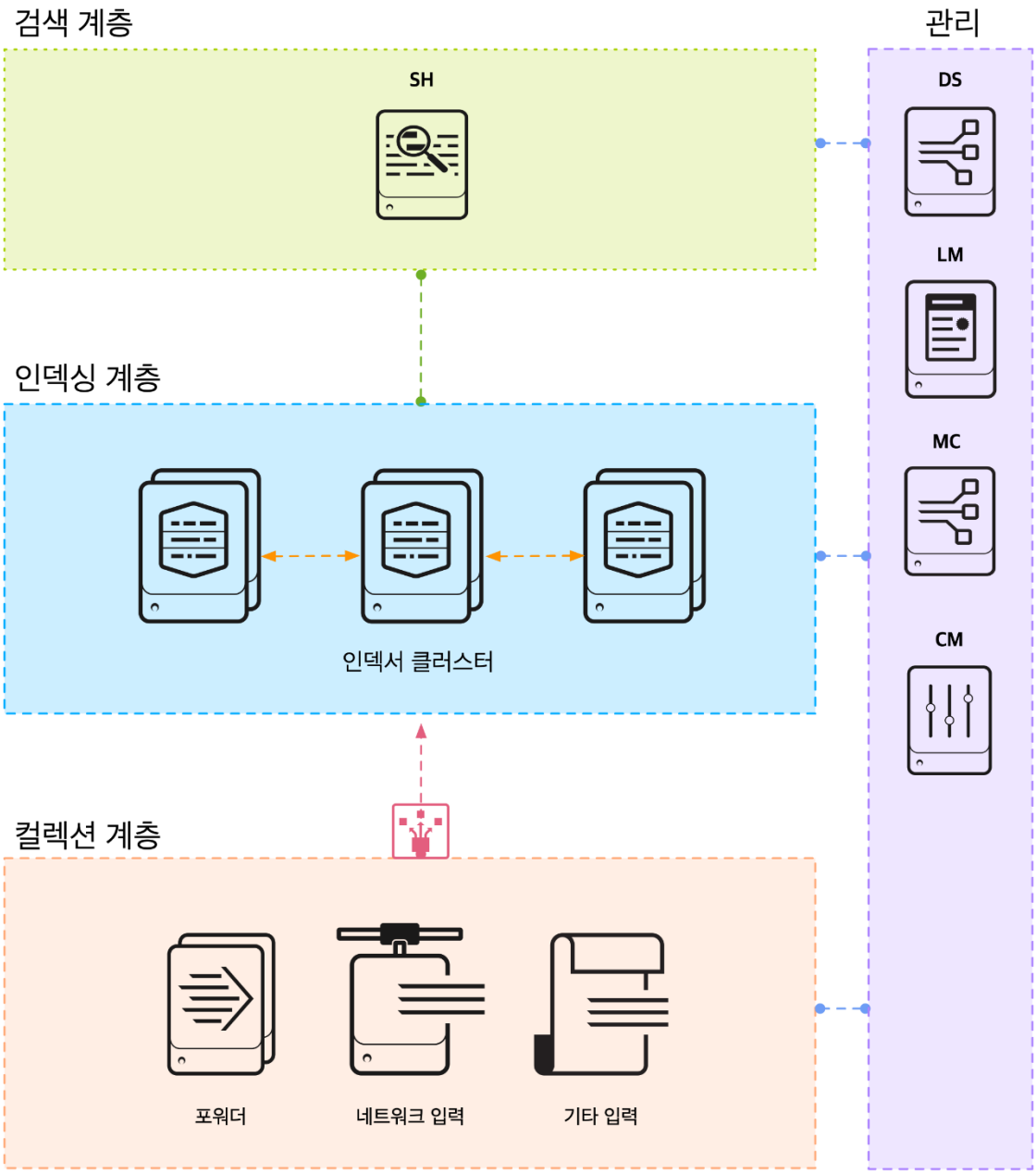
클러스터된 배포 옵션

아래에서 다음과 같은 토폴로지 옵션을 찾을 수 있습니다.

배포 유형	토폴로지 범주 코드
클러스터된 분산 배포 - 단일 사이트	C1/C11
클러스터된 분산 배포 + SHC - 단일 사이트	C3/C13
클러스터된 분산 배포 - 멀티 사이트	M2/M12
클러스터된 분산 배포 + SHC - 멀티 사이트	M3/M13
클러스터된 분산 배포 + SHC - 멀티 사이트	M4/M14

토폴로지 구성 요소에 대한 설명은 아래의 부록 "B"를 참조하십시오.

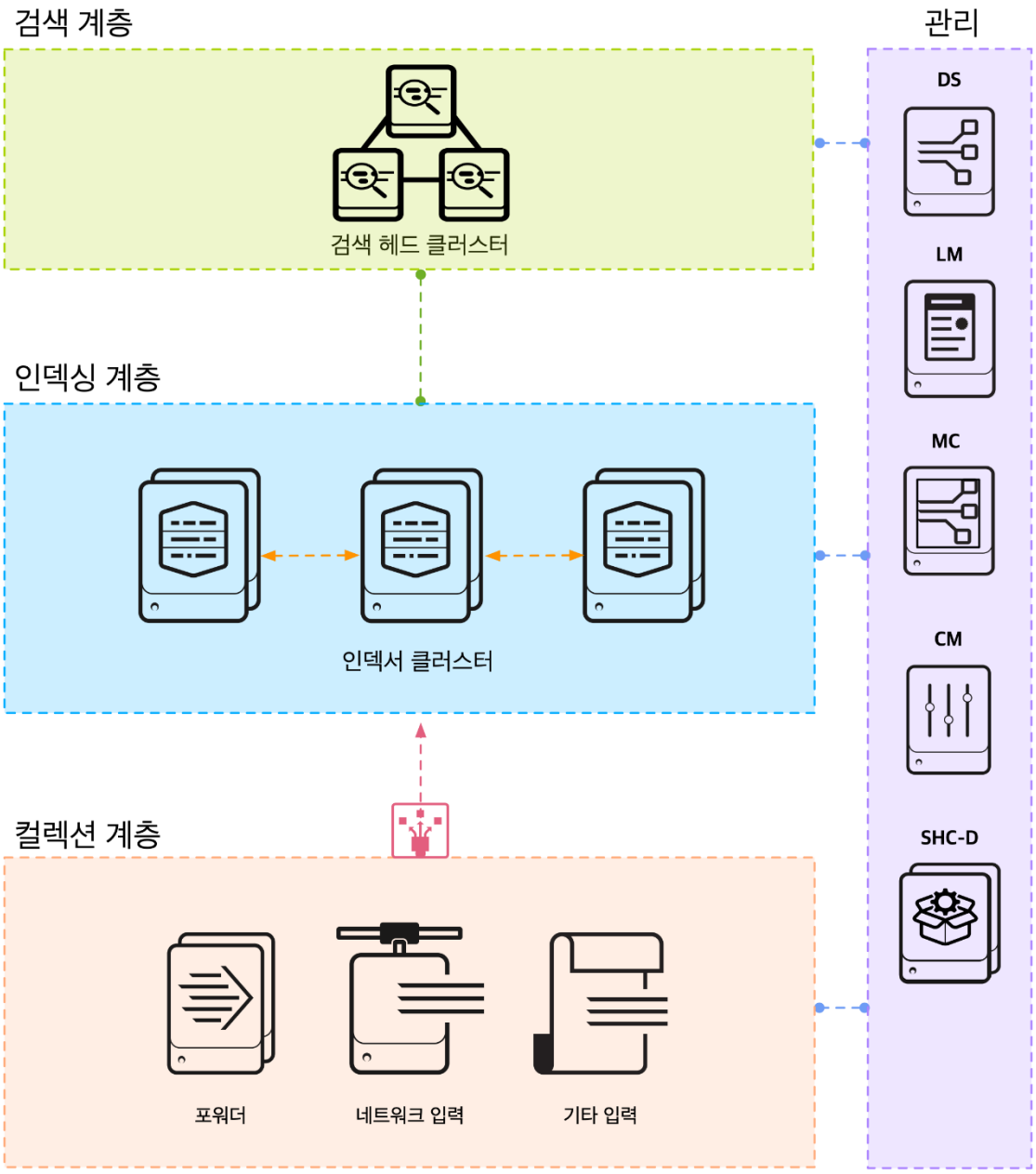
클러스터된 분산 배포 - 단일 사이트(C1/C11)



클러스터된 분산 배포에 대한 설명 - 단일 사이트(C1/C11)	제한 사항
이 토폴로지는 적절하게 설정된 데이터 복제 정책과 함께 인덱서 클러스터링을 도입합니다. 이렇게 하면 인덱서 피어 노드 오류의 경우 데이터의 고가용성을 제공할 수 있습니다. 하지만 인덱싱 계층에만 적용되며 검색 헤드 오류를 방지할 수는 없습니다. **ES 고객 참고 사항:** 범주 코드가 C11(예: Splunk App for Enterprise Security를 배포하려는 경우)인 경우 **전용 단일** 검색 헤드는 앱을 배포하는 데 필요합니다(토폴로지 다이어그램에는 표시되지 않음).	• 검색 계층을 위한 고가용성 없음 • 인덱서 클러스터의 총 고유 버킷 수는 5MM(V6.6 이상), 총 버킷 수는 15MM임 • 데이터 센터 작동 중단 시 자동 DR 기능 없음

클러스터된 분산 배포에 대한 설명 - 단일 사이트(C1/C11)	제한 사항
이 토폴로지에는 Cluster Master(CM)라는 추가 Splunk 구성 요소가 필요합니다. CM은 구성된 데이터 복제 정책의 조정 및 시행을 담당합니다. 또한 CM은 사용 가능한 클러스터 피어(인덱서)의 신뢰할 수 있는 원본입니다. 검색 헤드 설정은 개별 검색 피어 대신 CM을 설정하여 간소화됩니다. CM을 통해 사용 가능한 인덱서를 검색하도록 전달 계층을 설정할 수 있습니다. 이렇게 하면 전달 계층을 간편하게 관리할 수 있습니다. 데이터는 클러스터 내에서 비결정적인 방식으로 복제됩니다. 이에 따라 각 이벤트의 요청된 사본이 저장되는 위치를 제어할 수 없습니다. 또한 확장성은 선형적이지만 전체 클러스터 크기(이상적인 조건에서 검색 가능한 데이터의 ~50PB)와 관련된 제한 사항이 있습니다. Splunk 환경의 상태를 모니터링하기 위해 Monitoring Console(MC)을 배포하는 것이 좋습니다.	

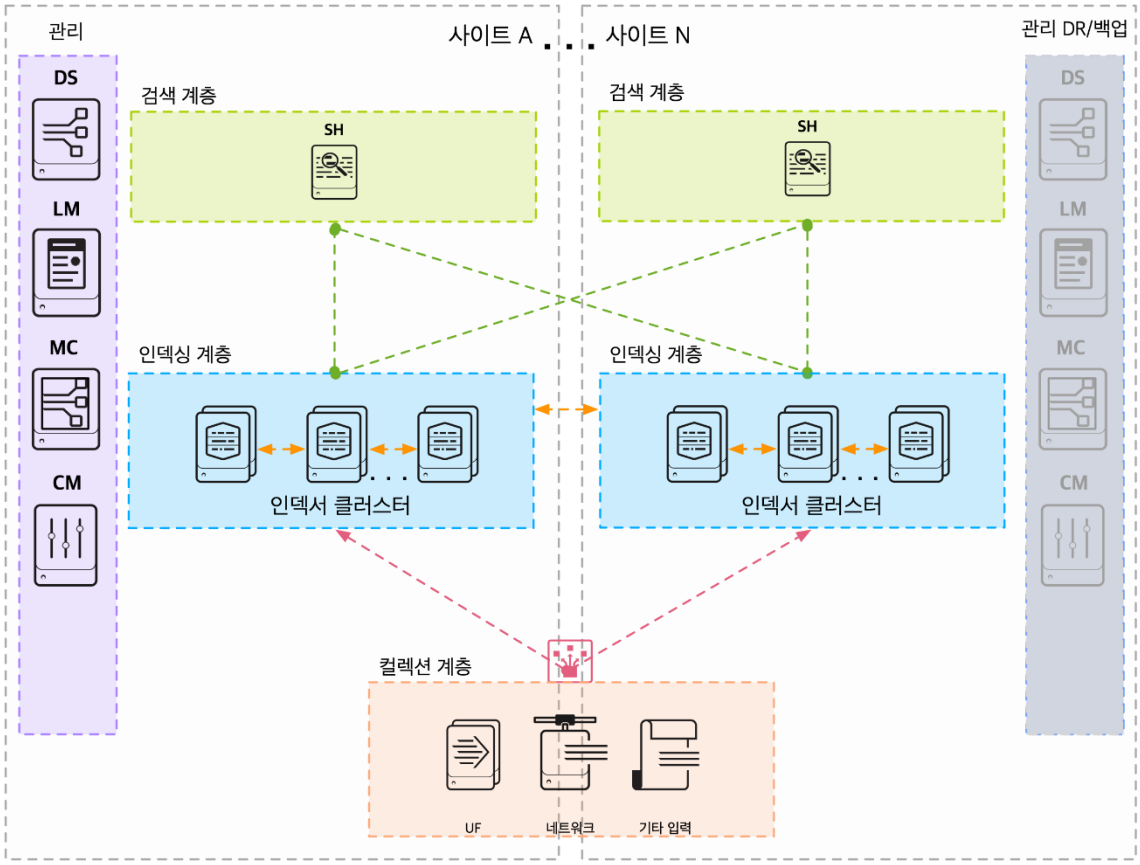
클러스터된 분산 배포 + SHC - 단일 사이트(C3/C13)



클러스터된 분산 배포에 대한 설명 + SHC - 단일 사이트(C3/C13)	제한 사항
이 토폴로지는 수평 확장성을 추가하고 검색 계층에서 단일 오류 지점을 제거합니다. SHC를 구현하려면 검색 헤드가 3개 이상 필요합니다. SHC 설정을 관리하려면 Search Head Cluster Deployer라는 추가 Splunk 구성 요소가 각 SHC에 필요합니다. 이 구성 요소는 클러스터의 설정 파일에 대한 변경 사항을 배포하는 데 필요합니다. Search Head Cluster Deployer에는 HA 요구 사항이 없습니다(런타임 역할 없음). SHC는 단일 검색 헤드가 제공할 수 있는 용량 이상으로 사용 가능한 검색 용량을 늘릴 수 있는 메커니즘을 제공합니다. 또한 SHC를	• 데이터 센터 작동 중단 시 DR 기능 없음 • ES에는 전용 SH/SHC가 필요함 • SHC에서 ES 배포 관리가 지원되지만

클러스터된 분산 배포에 대한 설명 + SHC - 단일 사이트(C3/C13)	제한 사항
사용하면 클러스터 전체에 예약된 검색 워크로드를 분산할 수 있습니다. SHC는 검색 헤드에 오류가 발생할 경우 최적의 사용자 페일오버를 제공합니다. 고정 세션을 지원하는 네트워크 부하 분산 장치는 SHC 구성원 앞에 있어야 클러스터 전체에 걸쳐 사용자 부하 분산이 적절하게 유지됩니다. **ES 고객 참고 사항:** 범주 코드가 C13(예: Splunk App for Enterprise Security를 배포하려는 경우)인 경우 **전용** 검색 헤드 클러스터는 앱을 배포하는 데 필요합니다(토폴로지 다이어그램에는 표시되지 않음). 검색 계층에는 사용자의 용량 및 조직 요구 사항에 따라 클러스터된 검색 헤드와 클러스터되지 않은 검색 헤드가 포함될 수 있습니다(이는 토폴로지 다이어그램에는 표시되지 않음).	문제가 있음(PS 포함) • SHC는 100개 이상의 노드를 가질 수 없음

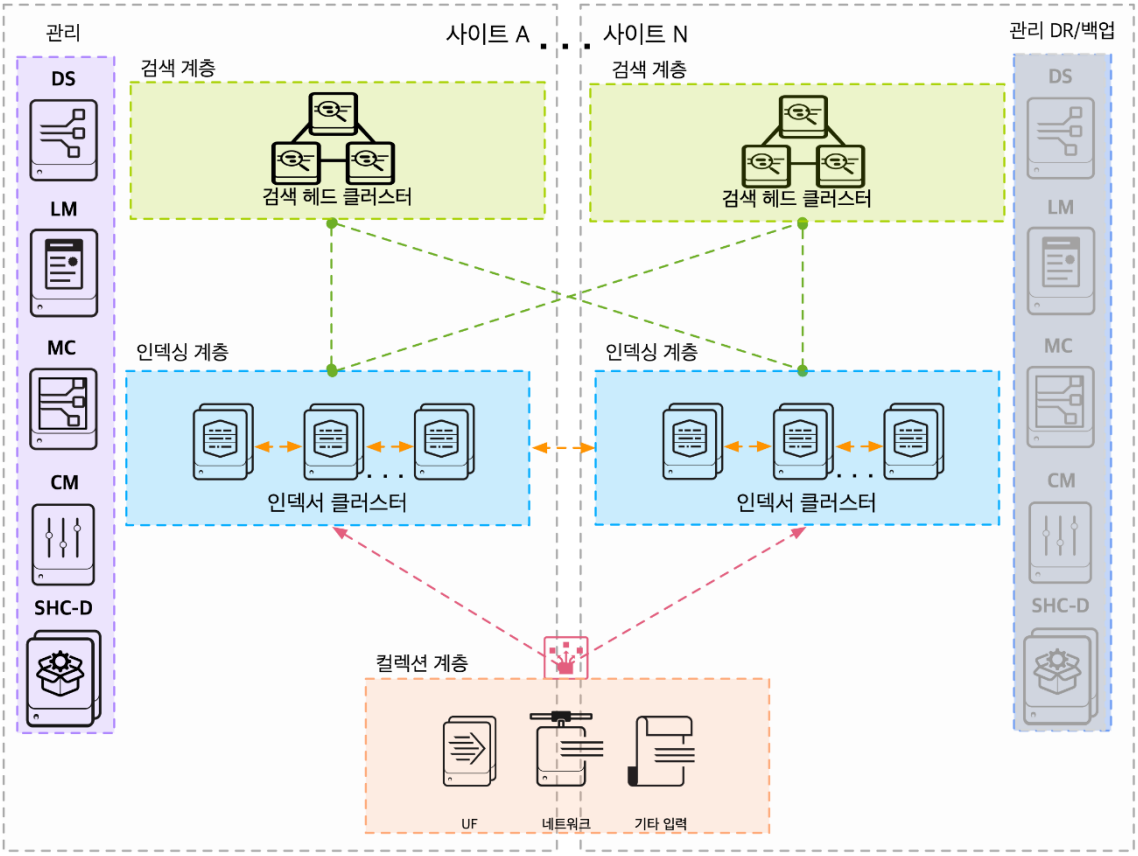
클러스터된 분산 배포 - 멀티 사이트(M2/M12)



클러스터된 분산 배포에 대한 설명 - 멀티 사이트(M2/M12)	제한 사항
치명적인 이벤트(예: 데이터 센터 작동 중단)가 발생할 경우 거의 자동식 재해 복구를 제공하기 위해 멀티 사이트 클러스터링이 가장 적합한 배포 아키텍처입니다. 성능이 좋은 멀티 사이트 클러스터는	• 사용 가능한 검색 헤드 용량을 공유하지 않고 **사이트에서** 검색

클러스터된 분산 배포에 대한 설명 - 멀티 사이트(M2/M12)	제한 사항
Splunk 매뉴얼에 지정된 대로 허용 가능한 사이트 간 네트워크 대기 시간이 필요합니다. 이 토폴로지를 사용하면 두 개 이상의 인덱서 클러스터 피어 그룹에 데이터를 **문제없이** 복제할 수 있습니다. 사이트 복제 및 검색 팩터를 설정할 수 있습니다. 이 사이트 복제 팩터를 사용하면 복제 복사본을 전송할 위치를 지정할 수 있으며 데이터가 여러 위치에 분산되도록 할 수 있습니다. 재해 발생 시 DR 사이트로 페일오버해야 하는 단일 클러스터 마스터 노드에서 여전히 관리합니다. 멀티 사이트 클러스터링은 물리적으로 별도로 분산된 위치에서 데이터 중복성을 제공하며 지리적으로 분산되어 배포될 수 있습니다. 사용자는 가용성을 보장하기 위해 DR 사이트에 자동으로 페일오버할 수 있습니다. 하지만 이 토폴로지는 검색 계층 설정과 런타임 아티팩트를 사이트 간에 자동으로 동기화하는 메커니즘을 제공하지 않습니다. 전체 사이트에 걸쳐 사용 가능한 검색 피어(인덱서) 용량을 활성/활성 모델에서 검색을 실행하는 데 활용할 수 있습니다. 가능할 경우 특정 사이트의 검색 헤드에 로그온한 사용자가 로컬 인덱서만 검색하도록 사이트 유사성을 설정할 수 있습니다. **ES 고객 참고 사항:** 범주 코드가 M12(예: Splunk App for Enterprise Security를 배포하려는 경우)인 경우 **전용 단일** 검색 헤드는 앱을 배포하는 데 필요합니다(토폴로지 다이어그램에는 표시되지 않음). ES 검색 헤드의 경우 페일오버에는 DR 상황에서만 활성화되고 사용되는 페일오버 사이트에서 "섀도" 검색 헤드를 설정해야 합니다. Splunk 프로페셔널 서비스를 통해 Enterprise Security 배포를 위한 사이트 페일오버 메커니즘을 설계하고 구현하십시오.	아티팩트 복제를 사용하지 않습니다. • 사이트 장애 발생 시 Splunk 외부에서 관리 기능의 오류를 처리해야 합니다. • 인덱스 복제에 대한 사이트 간 대기 시간은 권장 제한 범위 내에 있어야 합니다.

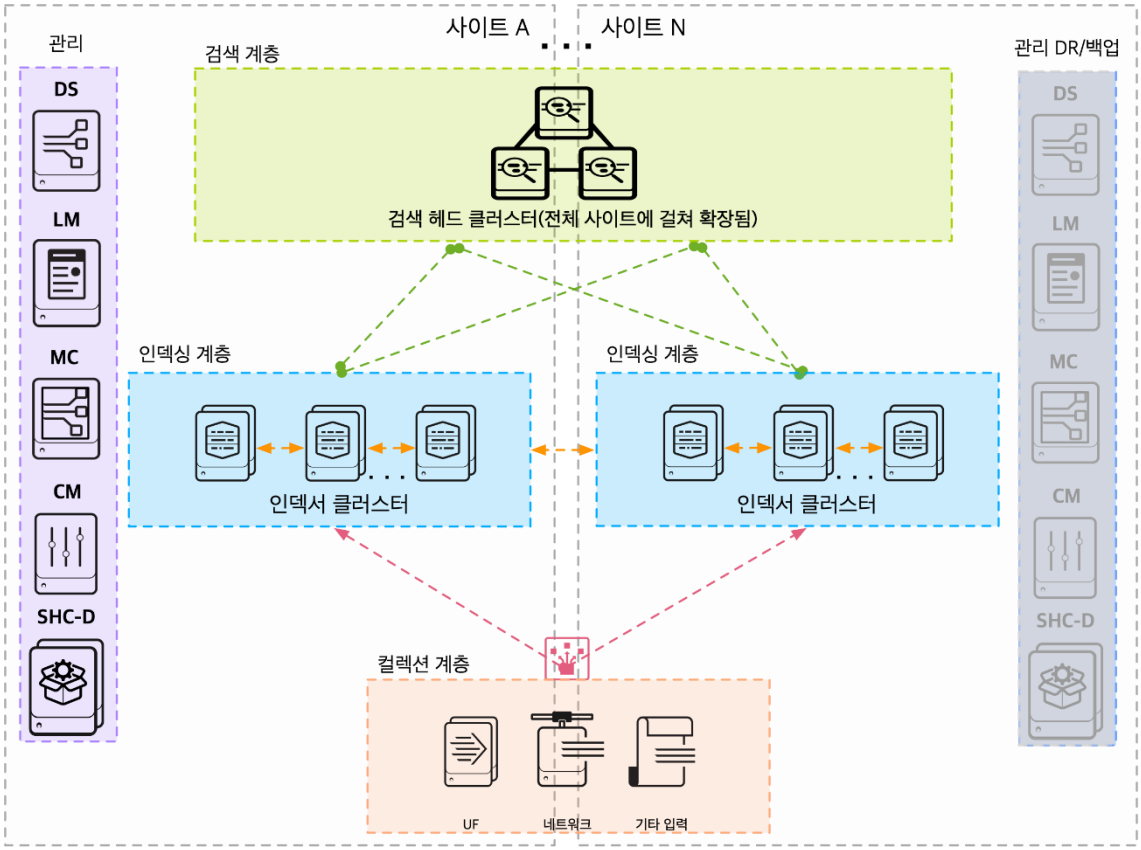
클러스터된 분산 배포 + SHC - 멀티 사이트(M3/M13)



클러스터된 분산 배포에 대한 설명 + SHC - 멀티 사이트(M3/M13)	제한 사항
이 토폴로지는 수평 확장성을 추가하고 각 사이트의 검색 계층에서 단일 오류 지점을 제거합니다. SHC(사이트별)를 구현하려면 검색 헤드가 3개 이상 필요합니다. SHC 설정을 관리하려면 Search Head Cluster Deployer라는 추가 Splunk 구성 요소가 **각** SHC에 필요합니다. 이 구성 요소는 클러스터의 설정 파일에 대한 변경 사항을 배포하는 데 필요합니다. Search Head Cluster Deployer에는 HA 요구 사항이 없습니다(런타임 역할 없음). SHC는 다음과 같은 이점을 제공합니다. a) 단일 검색 헤드가 제공할 수 있는 이점 이상으로 사용 가능한 검색 용량 증가, b) 클러스터 전체에 걸쳐 예약 검색 워크로드 분산, c) 검색 헤드 오류 발생 시 최적의 사용자 페일오버. 고정 세션을 지원하는 네트워크 부하 분산 장치는 각 사이트의 SHC 구성원 앞에 있어야 클러스터 전체에 걸쳐 사용자 부하 분산이 적절하게 유지됩니다.	• 사이트 전체에 걸쳐 검색 아티팩트 복제가 없으며 SHC는 독립형입니다. • 인덱스 복제에 대한 사이트 간 대기 시간은 문서로 지정된 제한 범위 내에 있어야 합니다. • SHC는 100개 이상의 노드를 가질 수 없음

클러스터된 분산 배포에 대한 설명 + SHC - 멀티 사이트(M3/M13)	제한 사항
ES 고객 참고 사항: 범주 코드가 M13(예: Splunk App for Enterprise Security를 배포하려는 경우)인 경우 사이트에 포함된 전용 단일 검색 헤드 클러스터는 앱을 배포하는 데 필요합니다(토폴로지 다이어그램에는 명확히 표시되지 않음). 사이트 장애로부터 ES SH 환경을 복구할 수 있도록 타사 기술을 사용하여 검색 헤드 인스턴스의 페일오버를 수행하거나 "웜 대기" ES SH를 프로비저닝하여 기본 ES 환경과 동기화할 수 있습니다. HA/DR 환경에서 ES를 배포할 경우 Splunk 프로페셔널 서비스에 참여하는 것이 좋습니다.	

클러스터된 분산 배포 + SHC - 멀티 사이트(M4/M14)



클러스터된 분산 배포에 대한 설명 + SHC - 멀티 사이트(M4/M14)	제한 사항
이 아키텍처는 고가용성 및 재해 복구에 대한 엄격한 요구 사항이 있는 배포용으로 설계된 가장 복잡한 검증된 아키텍처입니다. 적절한 배포를 위해 Splunk 프로페셔널 서비스를 사용하는 것이 좋습니다. 이 토폴로지를 적절히 배치하면 데이터 컬렉션, 인덱싱 및 검색을 위해 Splunk 인프라를 지속적으로 운영할 수 있습니다.	• 전체 사이트에 걸친 네트워크 대기 시간은 문서로 지정된 제한 범위 내에 있어야 합니다.

클러스터된 분산 배포에 대한 설명 + SHC - 멀티 사이트(M4/M14)	제한 사항
이 토폴로지에는 하나 이상의 사이트에 걸쳐 "확장된" 검색 헤드 클러스터 구현이 포함됩니다. 이는 검색 노드나 데이터 센터에 오류가 발생할 경우 사용자에게 최적의 페일오버를 제공합니다. 검색 아티팩트 및 기타 런타임 knowledge objects가 SHC에 복제됩니다. SHC 자체가 사이트를 인식하지 못하므로 (예: 아티팩트 복제는 비결정적임) 사이트 전체에서 복제되도록 주의 깊게 설정해야 합니다. 검색이 로컬에서 충족될 수 없는 경우에만 사이트 간의 WAN 링크를 활용하도록 사이트 유사성을 설정할 수 있습니다. 고정 세션을 지원하는 네트워크 부하 분산 장치는 SHC 구성원 앞에 있어야 클러스터 전체에 걸쳐 사용자 부하 분산이 적절하게 유지됩니다. **ES 고객 참고 사항:** 범주 코드가 M14(예: Splunk App for Enterprise Security를 배포하려는 경우)인 경우 사이트에 포함된 전용 단일 검색 헤드 클러스터는 앱을 배포하는 데 필요합니다(토폴로지 다이어그램에는 명확히 표시되지 않음). ES는 일관성 있는 런타임 아티팩트 집합을 사용할 수 있어야 하며 사이트 작동 중단이 발생할 경우 확장된 SHC에서 이를 보장할 수 없습니다. 사이트 장애로부터 ES SH 환경을 복구할 수 있도록 타사 기술을 사용하여 검색 헤드 인스턴스의 페일오버를 수행하거나 "웜 대기" ES SH를 프로비저닝하여 기본 ES 환경과 동기화할 수 있습니다. HA/DR 환경에서 ES를 배포할 경우 Splunk 프로페셔널 서비스에 참여하는 것이 좋습니다.	• 소수의 클러스터 구성원만 남은 경우 SHC의 페일오버 시 수작업을 수행해야 할 수 있습니다.

1b단계: 데이터 컬렉션 요구 사항 정의

데이터 컬렉션 계층은 Splunk 배포의 핵심 구성 요소입니다. 사용자 환경의 모든 장치가 데이터를 인덱싱 계층으로 전달하여 처리할 수 있기 때문에 Splunk에서 검색할 수 있습니다. Splunk 배포의 성공적인 구현 여부와 성능에 매우 중요한 요소는 전달 및 인덱싱 작업이 가장 효율적이고 안정적인 방식으로 이루어지도록 보장하는 것입니다.

데이터 컬렉션 계층 아키텍처에 대해 다음 측면을 고려하십시오.

- 데이터의 출처. 로그 파일, syslog source, 네트워크 입력 정보, OS 이벤트 로깅 기능, 애플리케이션, 메시지 버스 또는 다른 위치에서 데이터가 생성됩니까?
- 데이터 수집 대기 시간 및 처리량 요구 사항
- 인덱싱 계층의 인덱서 전반에 걸쳐 이상적인 이벤트 분산

- 오류 허용 및 자동 복구(HA)
- 보안 및 데이터 주권 요구 사항

이 SVA 절에서는 일반적인 데이터 컬렉션 방법을 중점적으로 다룹니다. 또한 이 절에서는 각 데이터 컬렉션 방법에 대한 아키텍처와 모범 사례에 대해 설명하고 구현 선택 시 고려해야 할 잠재적인 문제를 호출합니다.

주요 아키텍처 고려 사항과 이 고려 사항이 중요한 이유

데이터 컬렉션 계층의 필수적인 역할을 감안할 때 아키텍처 설계와 관련된 주요 고려 사항을 이해하는 것이 중요합니다.

이러한 고려 사항 중 일부는 사용자의 요구 사항에 따라 사용자와 관련이 있거나 관련이 없을 수도 있습니다. 하지만 아래 테이블의 굵게 표시된 텍스트는 모든 환경과 관련된 기본 항목을 설명합니다.

고려 사항	**이 고려 사항이 중요한 이유는 무엇입니까?**
데이터가 올바르게 수집됨(타임스탬프, 줄 바꿈, 잘림)	전체 인덱서에 걸쳐 이상적인 이벤트 분산 방식을 구현하는 것이 매우 중요합니다. 인덱싱 계층은 사용 가능한 모든 인덱서가 균등하게 활용될 때 가장 효율적으로 작동합니다. 이는 데이터 수집 및 검색 성능 모두에 해당됩니다. 피어에 비해 훨씬 많은 데이터를 처리하는 단일 인덱서는 검색 응답 시간에 부정적인 영향을 미칠 수 있습니다. 제한된 로컬 디스크 스토리지가 있는 인덱서의 경우 균등하지 않은 이벤트 분산으로 인해 설정된 데이터 보존 정책을 충족하기 전에 데이터가 조기에 삭제(age out)될 수 있습니다.
사용 가능한 인덱서에 데이터가 최적으로 분산됨	이벤트 타임스탬프 및 줄 바꿈이 올바르게 설정되지 않아 데이터가 제대로 수집되지 않을 경우 이 데이터를 검색하는 것이 매우 어려워집니다. 이는 검색 시간에 이벤트 경계를 강제로 시행해야 하기 때문입니다. 타임스탬프 추출 설정이 올바르지 않거나 누락되면 원하지 않는 암시적 타임스탬프가 할당될 수 있습니다. 이렇게 하면 사용자가 혼란스러울 수 있으며 데이터에서 가치를 얻기 훨씬 더 어려워집니다.
모든 데이터가 손실 없이 안정적으로 인덱싱 계층에 도달함	신뢰할 수 있는 분석을 위해 수집된 로그 데이터가 완전하고 유효해야 해당 데이터를 사용하여 수행된 검색이 유효하고 정확한 결과를 제공할 수 있습니다.
모든 데이터가 최소 대기 시간으로 인덱싱 계층에 도달함	데이터 수집 지연은 잠재적으로 치명적인 이벤트 발생과 해당 이벤트에 대한 검색 및 대응 기능 간에 소요되는 시간을 늘립니다.

	최소 수집 대기 시간은 직원에게 경고를 보내거나 자동화된 조치를 취하는 이용 사례를 모니터링하는 데 중요합니다.
데이터가 안전하게 전송됨	데이터가 민감하거나 데이터를 신뢰할 수 없는 네트워크를 통해 전송하는 동안 보호해야 하는 경우 타사의 무단 차단을 방지하기 위해 데이터를 암호화해야 할 수 있습니다. 일반적으로 SSL을 활성화하려면 Splunk 구성 요소 간의 모든 연결을 사용하는 것이 좋습니다.
네트워크 리소스 사용이 최소화됨	로그 데이터 컬렉션에 미치는 네트워크 리소스 영향을 최소화하여 기타 비즈니스 크리티컬 네트워크 트래픽에 영향을 미치지 않도록 해야 합니다. 전용 회선 네트워크의 경우 네트워크 사용을 최소화하면 배포 시 TCO를 절감할 수도 있습니다.
데이터 원본 인증/승인	불량 데이터 원본이 인덱싱 환경에 영향을 미치지 않도록 하려면 연결 인증/승인을 구현하는 것이 좋습니다. 이렇게 하면 네트워크 제어를 사용하거나 애플리케이션 레벨 메커니즘(예: SSL/TLS)을 사용하여 처리할 수 있습니다.

이 매뉴얼의 지침에서는 배포 시 매우 중요한 역할을 하며 이상적인 이벤트 분산을 지원하는 아키텍처를 중점적으로 다룹니다. Splunk 환경이 기대되는 검색 성능을 제공하지 않을 때 거의 모든 경우가 최소 스토리지 성능 요구 사항을 충족하지 못하거나 검색 병렬 처리를 제한하는 불균등한 이벤트 분산으로 인해 발생합니다.

가장 중요한 아키텍처 고려 사항에 대해 알아보았습니다. 이제 사용자가 수행해야 할 특정 데이터 컬렉션 요구 사항에 대해 알아보겠습니다.

설문지 2: 데이터 컬렉션 요구 사항 정의

다음 질문에 답변하면 배포에 필요한 데이터 컬렉션 구성 요소 리스트가 제공됩니다. 맨 오른쪽 컬럼의 키를 사용하여 문서의 각 구성 요소에 대한 세부 정보를 확인할 수 있습니다.

#	질문	고려 사항	토폴로지에 미치는 영향	관련 데이터 컬렉션 구성 요소
1	엔드포인트에서 로컬 파일을 모니터링하거나 데이터 컬렉션 스크립트를 실행해야 합니까?	이는 거의 모든 Splunk 배포 시나리오의 핵심 요구 사항입니다.	엔드포인트에 유니버설 포워더를 설치하고 중앙에서 해당 설정을 관리해야 합니다.	UF

2	어플라이언스, 네트워크 스위치 등과 같은 소프트웨어를 설치할 수 없는 장치에서 syslog를 통해 전송된 로그 데이터를 수집해야 합니까?	Syslog는 사용자 지정 소프트웨어 설치를 허용하지 않는 특정 용도에 맞게 제작된 장치에서 자주 사용되는 유비쿼터스 전송 프로토콜입니다.	컬렉션 지점 역할을 하는 syslog 서버 인프라가 필요합니다.	SYSLOG HEC
3	로컬 디스크에 기록하지 않고 API에 기록하는 애플리케이션의 로그 데이터 컬렉션을 지원해야 합니까?	엔드포인트에서 로그 파일에 기록하려면 디스크 공간을 제공하고 이러한 로그 파일을 관리해야 합니다(순환, 삭제 등). 일부 고객의 경우 이 모델이 아닌 사용 가능한 로깅 라이브러리를 사용하여 Splunk에 직접 로그인하려고 합니다.	Splunk HTTP Event Collector(HEC) 또는 로그 싱크 역할을 하는 다른 기술을 사용해야 합니다.	HEC
4	스트리밍 이벤트 데이터 공급자로부터 데이터를 수집해야 합니까?	수많은 기업에서 AWS Kinesis 또는 Kafka 같은 중앙 집중식 스트리밍 데이터 플랫폼이 로그 데이터 생성자와 사용자 간의 메시지 전송 역할을 하는 이벤트 허브 모델을 도입했습니다.	스트리밍 데이터 공급자와 Splunk 간의 통합이 필요합니다.	KAFKA KINESIS HEC
5	로그 생성자가 인덱싱 계층과 직접 TCP 연결을 설정하지 못하게 하는 협상 불가능한 보안 정책이 있습니까?	네트워크 토폴로지는 제한된 방화벽 규칙이 있는 여러 네트워크 영역으로 구성되며 일반적으로 네트워크 영역 간에 Splunk 포트의 트래픽 흐름을 허용하지 않을 수 있습니다. 개별 원본/대상 IP 주소에 대한 방화벽 규칙을 설정하고	네트워크 영역 간의 트래픽 흐름을 허용하는 중간 전달 계층이 필요합니다.	IF

		유지하는 것은 번거로울 수 있습니다.		
6	프로그래밍 방식(예: REST API를 호출하거나 데이터베이스를 쿼리함)을 사용하여 로그 데이터를 수집해야 합니까?	Splunk는 관계형 데이터베이스에서 데이터를 수집하는 DBX를 포함하여 다양한 데이터 수집 이용 사례의 API에 대해 스크립트를 실행할 수 있는 다양한 모듈식 입력을 제공합니다.	데이터 컬렉션 계층에는 Splunk 헤비 포워더로 구현된 하나 이상의 데이터 컬렉션 노드(DCN)가 필요합니다.	DCN
7	Splunk 외에도 다른 시스템으로 데이터(하위 집합)를 라우팅해야 합니까?	일부 이용 사례에서는 Splunk에서 인덱싱된 데이터가 다른 시스템으로 전달되도록 요구합니다. 대개 전달된 데이터는 원본 데이터의 하위 집합으로만 구성되거나 전달되기 전에 데이터를 수정해야 합니다.	이용 사례의 세부 사항에 따라 이벤트 기반 라우팅 및 필터링을 지원하기 위해 헤비 포워더로 만든 중간 전달 계층이 필요할 수 있습니다. 또는 Splunk App for CEF에 포함된 cefout 명령어를 사용하여 데이터 사후 인덱싱을 전달할 수 있습니다.	HF
8	네트워크 대역폭 제약조건을 가진 원격 사이트가 있고 네트워크를 통해 전송되기 전에 중요한 데이터를 필터링해야 합니까?	데이터를 전송 전에 필터링하려면 파싱(헤비) 포워더가 필요합니다. HWF에서 사용되는 아웃바운드 네트워크 대역폭은 UF의 약 5배이며, 필터링을 통해 상당수의 이벤트가 필터링되면 필터링만 가능합니다 (경험에 입각한 기준: 원본 데이터의 50% 이상). 로그 볼륨을 줄이기 위해 로깅 세분화를 조정하는 것이 가장 좋습니다.	원본에서 로그 볼륨을 줄일 수 없는 경우 원격 사이트에서 원본 데이터를 파싱하고 설정을 기반으로 이벤트를 필터링하는 중개 HF가 필요합니다.	IF HF

9	중요한 데이터를 인덱싱하기 위해 공용 네트워크를 통해 전송하기 전에 마스킹/난독 처리해야 합니까?	SSL을 사용하여 포워더 트래픽을 보호하는 것은 공용 네트워크를 통해 전송되는 중요한 데이터를 보호하기에 충분한 방법이 아니므로, 전송하기 전에 이벤트의 각 부분을 마스킹해야 합니다(SSN, CC 데이터 등). 이상적으로 이러한 데이터 마스킹은 로그 데이터를 생성하는 애플리케이션에서 수행됩니다.	운영 애플리케이션에서 데이터를 마스킹할 수 없는 경우 원본 데이터를 파싱하고 인덱서에 데이터를 보내기 전에 설정을 기반으로 필요한 마스킹 규칙을 적용하는 중간 HF가 필요합니다.	IF HF
10	statsd 또는 collectd를 사용하여 메트릭을 캡처해야 합니까?	Statsd 및 collectd는 호스트 시스템 및 애플리케이션에서 메트릭을 수집하는 데 사용되는 유비쿼터스 기술입니다.	Splunk는 UF, HF 또는 HEC를 사용하여 이러한 인덱스를 공급하는 특정 인덱스 유형 및 컬렉션 방법을 지원합니다.	METRICS
11	고가용성을 위해 데이터 컬렉션 구성 요소가 필요합니까?	일반적으로 엔드포인트에는 적용할 수 없으며 중간 포워더 또는 데이터 컬렉션 노드 같은 기타 데이터 컬렉션 구성 요소의 가용성이 고려될 수 있습니다.	작동 중단이 각 구성 요소의 가용성에 어떻게 영향을 미치는지, 그리고 이를 어떻게 처리해야 하는지에 대해 고려해야 합니다.	HA

2b단계: 데이터 컬렉션 구성 요소 선택

설문지를 완료하면 배포 요구 사항을 충족하는 데 필요한 데이터 컬렉션 구성 요소 리스트가 제공됩니다. 이 절에서는 각 데이터 컬렉션 아키텍처 구성 요소에 대해 자세히 설명합니다. 직접 하기 전에 간단히 몇 가지 일반적인 지침을 제공합니다.

일반적인 전달 아키텍처 지침

데이터 컬렉션 계층은 가능한 한 "평평한" 것이 가장 이상적입니다. 즉 데이터 원본은 유니버설 포워더에서 로컬로 수집되어 인덱싱 계층으로 직접 전달됩니다. 이는 데이터 수집 대기

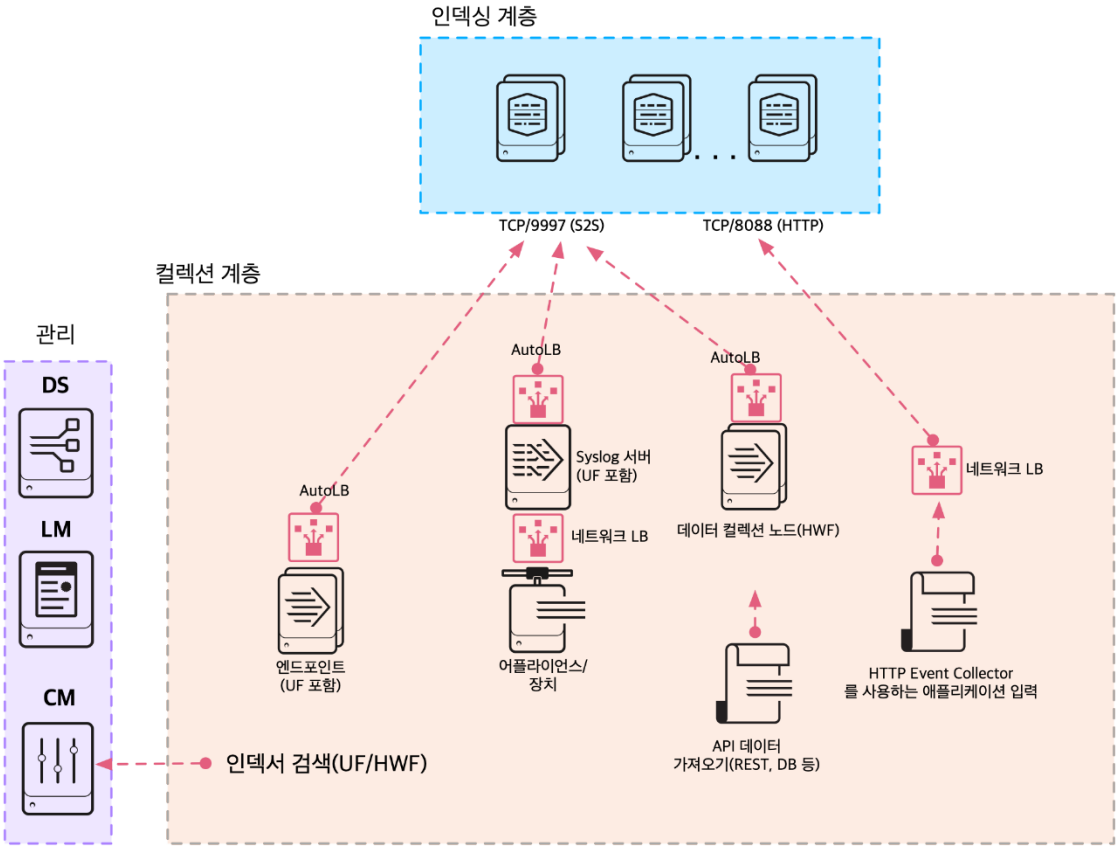
시간(검색 시간)을 최소화하고 사용 가능한 인덱서에서 적절한 이벤트 분산을 가능하게 하므로 최상의 방법입니다. 이 모범 사례를 따르면 관리가 쉬우며 운영이 간편합니다. 고객이 중간 전달 계층을 배포하는 경우도 있습니다. 일반적으로 요구 사항을 충족시킬 수 없는 경우를 제외하고 이 방식은 지양합니다. 중간 포워더의 잠재적인 영향으로 이 문서에는 이 항목에 대한 별도의 절과 세부 정보가 포함되어 있습니다.

유니버설 포워더 (즉, 네트워크 장치, 어플라이언스) 설치와 syslog 프로토콜을 사용하는 로그를 허용하지 않는 엔드포인트가 있습니다. 이러한 데이터 원본을 수집하는 별도의 모범 사례 아키텍처는 Syslog 데이터 컬렉션 절에 설명되어 있습니다.

프로그래밍 방식(API, 데이터베이스 액세스)을 사용하여 수집해야 하는 데이터 원본의 경우 전체 Splunk Enterprise 설치를 기반으로 하는 데이터 컬렉션 노드(DCN)를 배포하는 것이 좋습니다. 이 방식은 헤비 포워더로도 알려져 있습니다. 개발 환경이 아닌 다른 환경의 검색 헤드 계층에서 이러한 유형의 입력을 실행하지 않는 것이 좋습니다.

다음 다이어그램은 이 지침을 반영하는 일반적인 데이터 컬렉션 아키텍처를 보여줍니다.

데이터 컬렉션 토폴로지 개요



위의 다이어그램은 데이터 컬렉션 구성 요소의 설정을 관리하는 데 사용되는 관리 계층의 배포 서버(DS)를 보여줍니다. 또한 Splunk Enterprise 기능을 사용하려면 데이터 컬렉션 노드가 LM에 액세스해야 하므로 라이선스 마스터(LM)가 표시됩니다. 해당되는 경우 인덱서 검색 시

포워더에서 클러스터 마스터(CM)를 사용할 수 있으므로 포워더 출력 설정에서 사용 가능한 인덱서를 관리하지 않아도 됩니다.

위의 다이어그램에서 AutoLB는 Splunk에서 기본으로 제공되는 자동 부하 분산 메커니즘을 보여줍니다. 이 메커니즘은 Splunk 전용 S2S 프로토콜(기본 포트 9997)을 사용하여 전송된 데이터에 대해 적절한 이벤트 분산을 보장하는 데 사용됩니다. 참고: S2S 트래픽용 외부 네트워크 부하 분산기는 현재 지원되지 않으며 해당 분산기를 사용하도록 권장하지 않습니다.

업계 표준 프로토콜(예: HTTP 또는 syslog)을 통해 전송하는 데이터 원본의 부하 분산 트래픽의 경우 네트워크 부하 분산기를 사용하여 인덱싱 계층의 전체 인덱서에서 부하 및 이벤트 분산을 보장합니다.

유니버설 포워더(UF)

유니버설 포워더(UF)는 사용자 환경의 시스템에서 대량의 데이터 컬렉션 요구 사항에 가장 적합한 선택입니다. 이는 최소한의 리소스 요구 사항을 포함하는 특정 목적을 가진 데이터 컬렉션 메커니즘입니다. UF는 로그 데이터를 수집하고 전달하는 기본 선택 항목이어야 합니다. UF는 다음을 제공합니다.

- 무손실 데이터 컬렉션을 위한 체크포인트/다시 시작 기능
- 네트워크 대역폭 사용을 최소화하는 효율적인 프로토콜
- 조절 기능
- 사용 가능한 인덱서 전체에 부하 분산 기능이 기본 제공되어 있음
- SSL/TLS를 사용하는 선택적 네트워크 암호화
- 데이터 압축(SSL/TLS 없이 사용)
- 여러 입력 방식(파일, Windows 이벤트 로그, 네트워크 입력, 스크립트 기반 입력)
- 제한된 이벤트 필터링 기능(Windows 이벤트 로그 전용)
- 병렬식 수집 파이프라인 지원으로 처리량 증가/대기 시간 감소

적절하게 구조화된 데이터(json, csv, tsv)에 대한 예외는 거의 없습니다. 하지만 UF는 로그 원본을 이벤트로 파싱하지 않으므로 로그 형식을 파악하는 데 필요한 작업을 수행할 수 없습니다. 또한 Python을 제거한 버전과 함께 제공되므로 전체 Splunk 스택을 작동하는 데 필요한 모듈식 입력 앱과 호환되지 않습니다.

일반적으로 Splunk 환경의 엔드포인트와 서버에 다수의 UF(100 ~ 10,000s)를 배포하고 Splunk 배포 서버 또는 타사 설정 관리 도구(예: Puppet 또는 Chef)를 사용하여 중앙에서 관리됩니다.

헤비 포워더(HF)

헤비 포워더(HWF)는 인덱싱이 비활성화된 포워더의 역할을 하도록 설정된 전체 Splunk Enterprise 배포입니다. HWF는 일반적으로 다른 Splunk 역할을 수행하지 않습니다. HWF는 디스크에 이벤트를 실제로 작성 및 인덱싱하지 않고 인덱서의 수행 기능과 동일한 기능을

수행하는 동시에 전체 파싱 파이프라인을 포함한다는 점이 UF와 HWF의 주요 차이점입니다. 이를 통해 HWF는 개별 이벤트(예: 데이터를 마스킹하거나 이벤트 데이터를 기반으로 필터링 및 라우팅 수행)를 파악하고 이에 따른 조치를 취할 수 있습니다. 이 설치는 전체 Splunk Enterprise에 대한 설치이므로 데이터 컬렉션 작업에 제대로 작동하려면 전체 Python 스택이 필요하거나 Splunk HTTP Event Collector(HEC)의 엔드포인트 역할을 하는 모듈식 입력을 호스트할 수 있습니다. HWF는 다음과 같은 기능을 수행합니다.

- 데이터를 이벤트로 파싱
- 개별 이벤트 데이터를 기반으로 필터링 및 라우팅
- UF보다 더 큰 규모의 리소스 공간이 있음
- UF(~ 5x)보다 더 큰 네트워크 대역폭 범위가 있음
- 관리용 GUI

일반적으로 HWF는 데이터 컬렉션 목적으로 엔드포인트에 설치되지 않습니다. 대신 독립 실행형 시스템에서 데이터 컬렉션 노드(DCN) 또는 중간 전달 계층을 구현하는 데 사용됩니다. **다른 시스템의 데이터 컬렉션 요구 사항이 UF를 충족할 수 없는 경우에만 HWF를 사용합니다.**

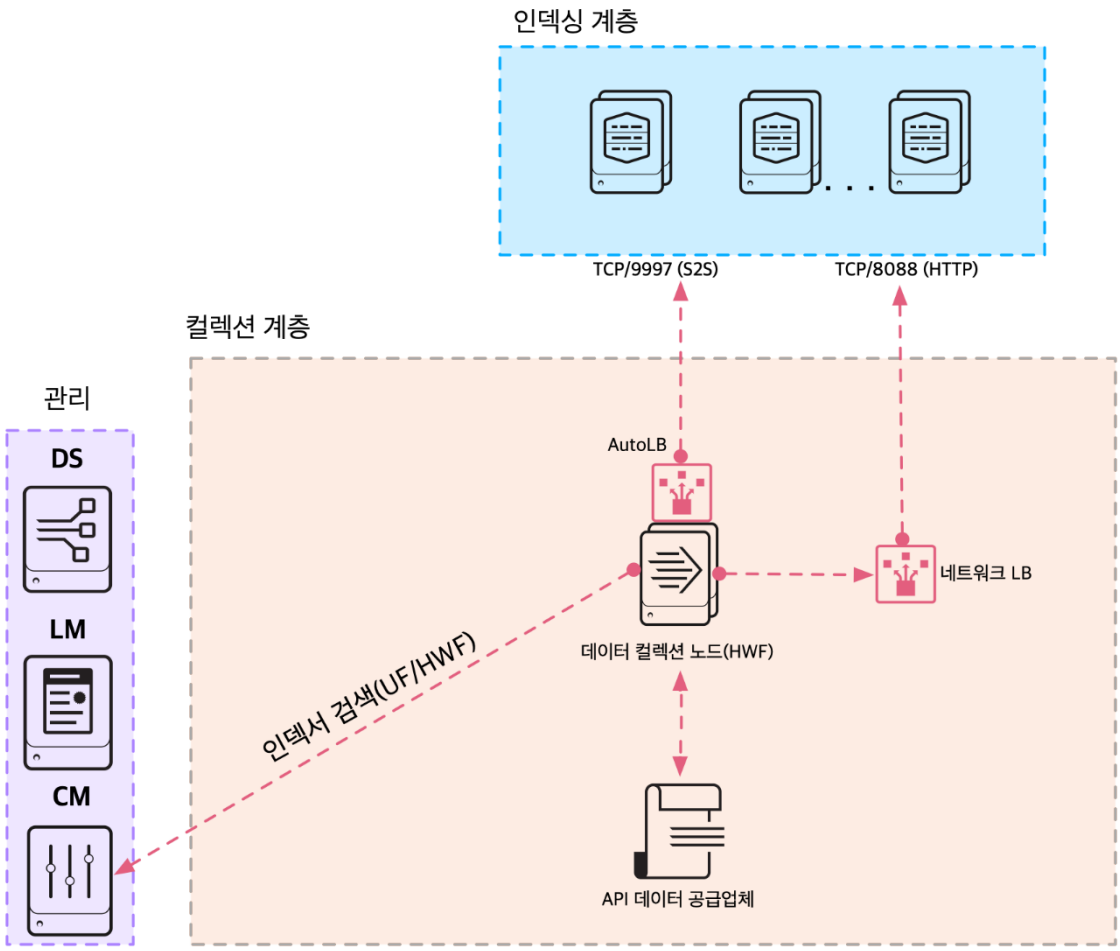
이러한 요구 사항의 예는 다음과 같습니다.

- Splunk로 데이터를 수집하기 위해 RDBMS에서 데이터를 읽음(데이터베이스 입력)
- API(클라우드 서비스, VMware 모니터링, 전용 시스템 등)를 통해 도달할 수 있는 시스템으로부터 데이터 수집
- HTTP Event Collector 서비스를 호스트하는 전용 계층 제공
- 라우팅/필터링/마스킹을 위한 파싱 포워더가 필요한 중간 전달 계층 구현

데이터 컬렉션 노드(DCN) 역할을 하는 헤비 포워더

일부 데이터 원본은 특정 종류의 API를 사용하여 수집해야 합니다. 이러한 API에는 쿼리 메커니즘으로 REST, 웹 서비스, JMS 및/또는 JDBC가 포함될 수 있습니다. 타사 개발자뿐 아니라 Splunk도 이러한 API 상호 작용을 허용하는 다양한 애플리케이션을 제공합니다. 일반적으로 이러한 애플리케이션은 Splunk 모듈식 입력 프레임워크를 사용하여 구현되며, 이 프레임워크가 제대로 작동하려면 전체 Splunk Enterprise 소프트웨어를 설치해야 합니다. 이 이용 사례를 실현하는 가장 좋은 방법은 데이터 컬렉션 노드(DCN)로 작동하도록 설정된 헤비 포워더로 작동하도록 하나 이상의 서버를 배포하는 것입니다.

데이터 컬렉션 노드 토폴로지

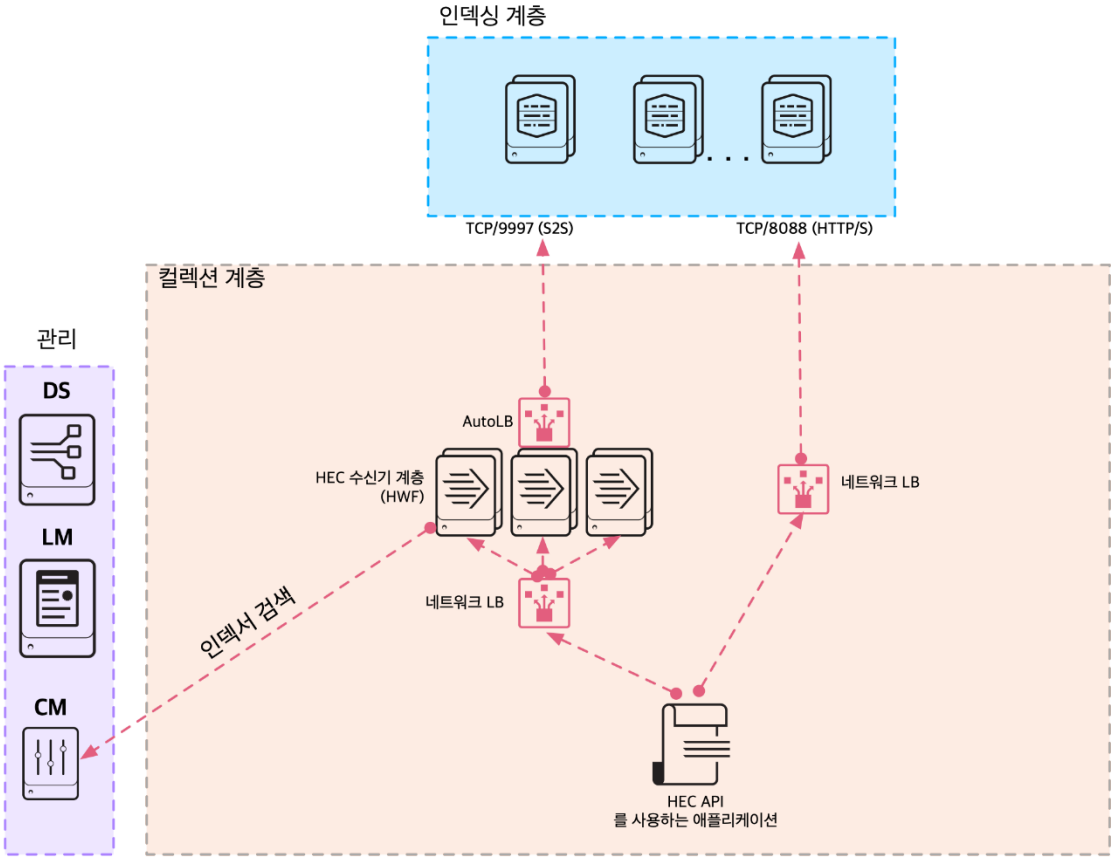


HTTP Event Collector(HEC)

HEC는 서버 측에서 HTTP/S 연결을 허용하는 수신기 서비스와 클라이언트 측에서 API를 제공하여 로그 데이터 페이로드를 인덱싱 계층이나 하나 이상의 헤비 포워더로 구성된 전용 HEC 수신기 계층에 직접 게시할 수 있도록 지원합니다. HEC는 원시 형식 또는 JSON 형식으로 전송할 데이터를 지원하는 두 개의 엔드포인트를 제공합니다. JSON을 활용하면 이벤트 페이로드에 추가 메타 데이터가 포함될 수 있으므로 향후 데이터를 검색할 경우 뛰어난 유연성을 제공합니다.

다음 다이어그램은 HEC의 두 가지 배포 옵션을 보여줍니다.

HEC 토폴로지 선택



관리 계층에는 라이선스 마스터(HF에서 필요함)뿐 아니라 수신되는 구성 요소에서 HTTP 입력을 관리하기 위한 배포 서버가 포함되어 있습니다. 참고: 인덱싱 계층이 클러스터되고 HEC 트래픽을 직접 수신하는 경우 HEC 설정은 배포 서버 대신 클러스터 마스터를 통해 관리됩니다.

어떤 배포 토폴로지를 선택할지는 대개 특정 요구에 따라 달라집니다. 전용 HEC 수신기 계층은 배포에 다른 아키텍처 구성 요소를 도입합니다. 긍정적 측면으로 보면 이는 독립적으로 확장할 수 있으며 관리 측면에서는 인덱싱 계층과 분리 레벨을 제공합니다. 또한 전용 HEC 계층에는 HF가 필요하므로 모든 인바운드 트래픽을 파싱하여 인덱서에서 워크로드를 제거합니다,

반면 인덱서에 직접 HEC 수신기를 호스트하면 HTTP가 모든 네트워크 부하 분산기에 적합한 프로토콜이므로 전체 인덱싱 계층에 걸쳐 한층 더 효과적으로 이벤트를 분산할 수 있으며, 적절한 부하 분산 정책을 통해 가장 부하가 적은 인덱서가 가장 먼저 제공하도록 보장할 수 있습니다.

요구 사항을 충족하는 가장 단순한 아키텍처를 배포한다는 면에서 HEC 수신기를 인덱서에 호스트하고, 호스트하기에 충분한 시스템 용량이 있다고 가정하는 것이 좋습니다. 적절한 크기로 설정된 HF 계층을 배포하고 인덱서 대신 HF의 IP 주소를 사용하도록 LB 설정을 변경하면 필요에 따라 추후에도 이 설정을 간편하게 변경할 수 있습니다. 이러한 변경 사항은 클라이언트 애플리케이션에 투명하게 처리되어야 합니다.

참고: HEC를 통해 전송된 데이터에 대해 인덱서 수신확인 기능을 요구할 경우 인덱서 다시 시작 롤링으로 인해 발생하는 중복 메시지를 최소화하려면 전용 HEC 수신기 계층을 사용하는 것이 좋습니다.

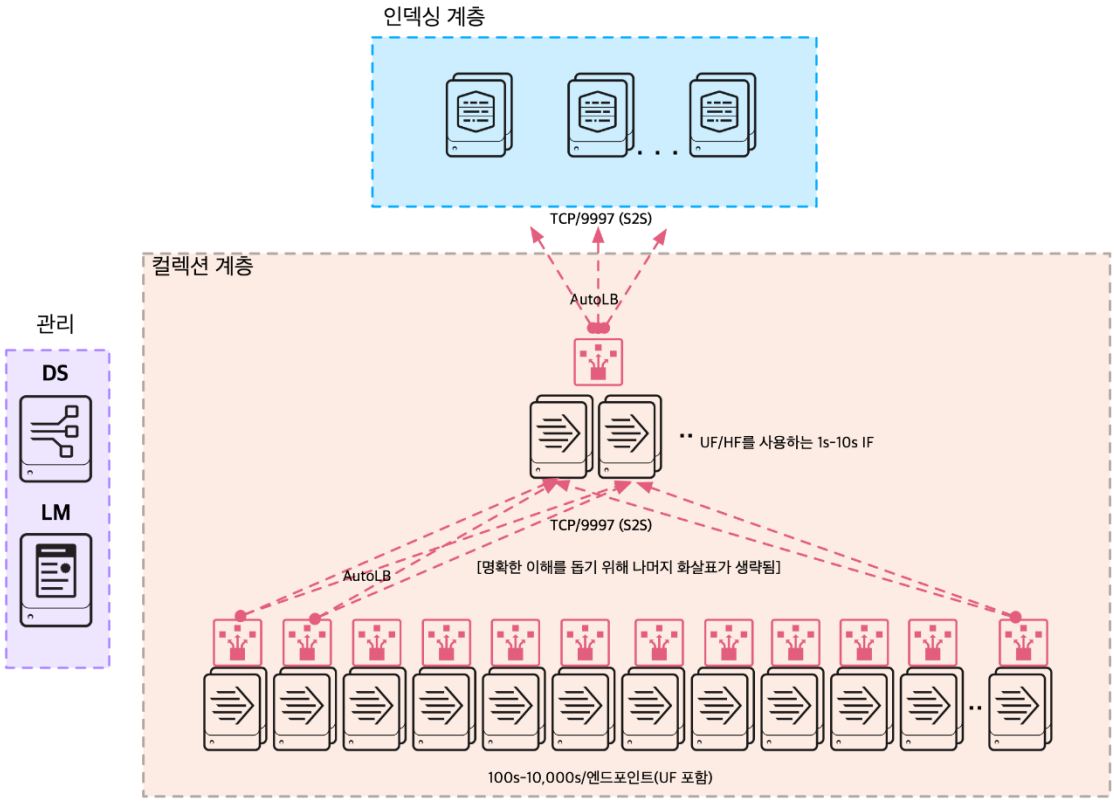
참고: 이 HEC 배포 아키텍처는 추후 논의할 일부 데이터 컬렉션 구성 요소, 특히 Syslog 및 메트릭 데이터 컬렉션에 대한 전송 기능을 제공합니다.

중간 전달 계층(IF)

경우에 따라 데이터 전달 시 중간 포워더가 필요합니다. 중간 포워더는 엔드포인트에서 로그 스트림을 수신하고 인덱서 계층으로 전달합니다. 중간 포워더는 전체 Splunk 환경에 부정적인 영향을 미치지 않도록 세심한 설계가 요구되는 아키텍처 과제를 안고 있습니다. 가장 중요한 것은 중간 포워더가 100 ~ 10,000명에 달하는 엔드포인트 포워더를 연결하고 훨씬 적은 수의 연결을 사용하여 인덱서로 전달한다는 점입니다. 이렇게 하면 인덱서의 하위 집합만 특정 시점에서 트래픽을 수신하므로 인덱싱 계층의 데이터 분산에 실질적으로 영향을 미칠 수 있습니다. 그러나 이러한 부작용은 적절한 규모 지정 및 설정으로 완화할 수 있습니다.

다음 다이어그램은 이러한 문제점을 잘 보여줍니다.

중간 전달 토폴로지



단일 중간 포워더가 있는 시나리오에서 모든 엔드포인트는 이러한 단일 포워더(잠재적으로 수천 개임)에 연결되며, 중간 포워더는 지정된 시간에 하나의 인덱서에만 연결됩니다. 이 시나리오는 다음과 같은 결과가 발생할 수 있으므로 최적의 시나리오는 아닙니다.

- 다수의 엔드포인트의 대규모 데이터 스트림은 시스템 및 네트워크 리소스를 많이 사용하는 단일 파이프를 통해 유입됩니다.
- 제한된 페일오버는 IF 오류의 경우 엔드포인트를 대상으로 수행됩니다(작동 중단 위험은 IF의 수에 반비례함).
- 소수의 인덱서가 특정 시점에서 제공됩니다. 단기간 동안 수행된 검색은 그 외의 경우만큼 병렬 처리의 다양한 이점을 누릴 수 없습니다.

중간 포워더는 추가적인 아키텍처 계층을 배포에 추가하므로 관리 및 문제 해결이 복잡해지고 데이터 수집 경로에 대기 시간이 추가됩니다. 요구 사항을 충족시키는 유일한 옵션이 아닐 경우 중간 전달 계층을 사용하지 마십시오. 다음과 같은 경우 중간 계층 사용을 고려해 볼 수 있습니다.

- 중요한 데이터의 경우 네트워크를 통해 인덱서로 전송하기 전에 난독화/제거해야 합니다. 예를 들어 공용 네트워크를 사용해야 하는 경우가 있습니다.
- 엄격한 보안 정책은 다중 영역 네트워크 또는 클라우드 기반 인덱서 같은 엔드포인트와 인덱서 간의 직접 연결을 허용하지 않습니다.
- 이벤트의 주요 하위 집합을 필터링해야 하는 엔드포인트와 인덱서 간의 대역폭 제약조건
- 동적 대상에 대한 이벤트 기반 라우팅이 요구 사항입니다.

이 계층의 가용성을 보장하고 모든 트래픽을 처리할 수 있는 충분한 처리 용량을 제공하며, 전체 인덱서에서 적절한 이벤트 분산을 지원하려면 모든 중간 전달 계층에 대한 규모 지정 및 설정 요구 사항을 고려하십시오. IF 계층에는 다음 요구 사항이 있습니다.

- 전반적으로 충분한 수의 데이터 처리 파이프라인이 있습니다.
- 중복 IF 인프라
- Splunk 부하 분산 설정이 올바르게 조정되었습니다. 예를 들어 autoLBVume, EVENT_BREAKER, EVENT_BREAKER_ENABLE, 가능할 경우 forceTimeBasedAutoLB가 있습니다(필요할 경우).

일반 지침의 경우 IF 처리 파이프라인을 인덱싱 계층에 있는 인덱서 양의 두 배로 하는 것이 좋습니다.

참고: 처리 파이프라인은 물리적 IF 서버와 동일하지 않으며, 충분한 시스템 리소스를 제공했습니다. 예를 들어 CPU 코어, 메모리 및 NIC 대역폭을 사용할 수 있는 경우 단일 IF를 여러 처리 파이프라인으로 설정할 수 있습니다.

IF 계층이 필요한 경우(설문지 참조), 기본적으로 계층에 UF를 사용하도록 설정되어 있으므로 시스템 및 네트워크에 대해 더 적은 리소스 저장 공간에서 더 높은 처리량을 제공합니다. UF의 기능이 요구 사항을 충족시키지 못할 경우 HF를 사용하십시오.

Syslog 데이터 컬렉션(SYSLOG)

Syslog 프로토콜은 기업 내 로그 데이터의 유비쿼터스 원본을 제공합니다. 확장성과 안정성이 가장 뛰어난 데이터 컬렉션 계층에는 syslog 수집 구성 요소가 포함되어 있습니다. Splunk로 syslog 데이터를 가져오는 여러 가지 방법이 있습니다. 다음과 같은 방법을 고려하십시오.

- 유니버설 포워더(UF)/헤비 포워더(HF) Splunk UF 또는 HF를 사용하여 syslog 서버가 작성한 파일(예: rsyslog 또는 syslog-ng)을 모니터링합니다.
- HEC에 대한 Syslog 에이전트: Splunk의 HEC로 출력할 수 있는 syslog 에이전트를 사용합니다. (HEC에 출력할 수 있는 rsyslog 및 syslog-ng용 타사 모듈이 있음) .
- **직접 TCP/UDP 입력**: Splunk는 TCP 또는 UDP 포트 (기본 포트는 UDP 514)에서 수신 대기하고 여기에 소스를 수집할 수 있습니다(프로덕션 용도로 권장되지 **않음**).

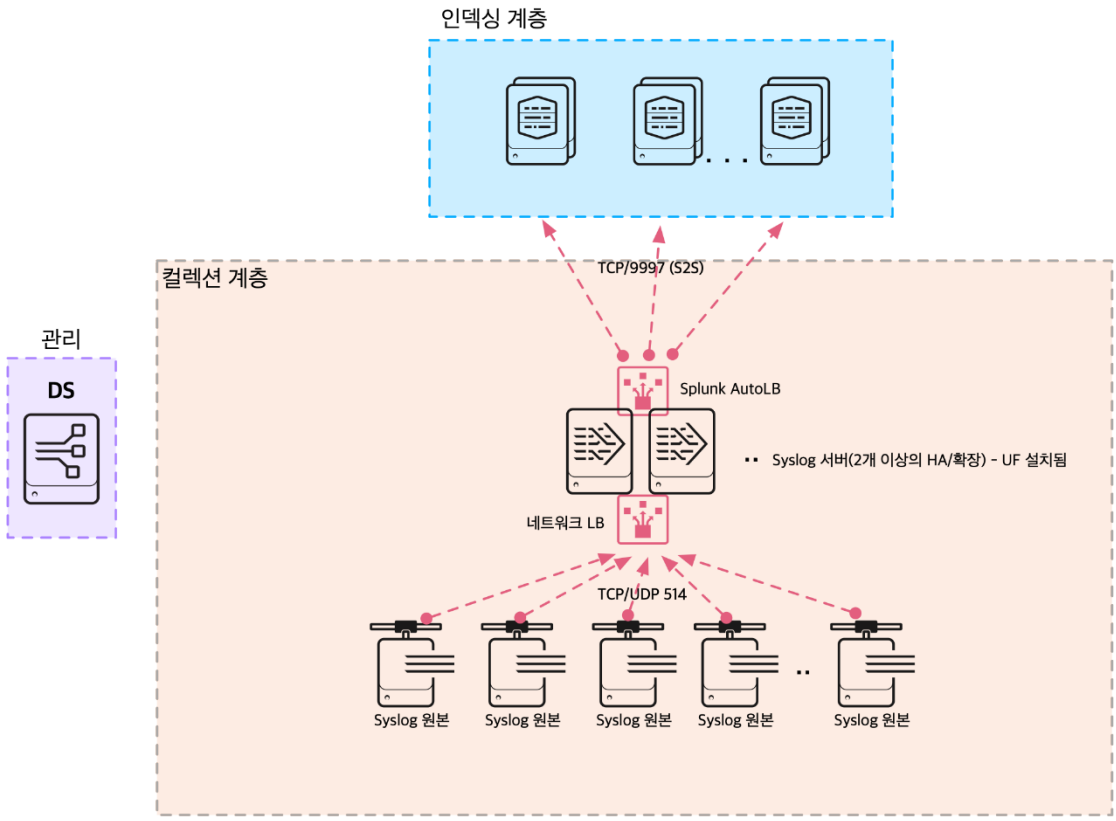
1.1.1.1 Syslog(SCD를 사용하여 파일 모니터링)

Splunk는 UF/HF에서 inputs.conf를 사용하여 모니터링하여 syslog 컬렉션 데몬(SCD)에서 엔드포인트의 디스크에 기록되는 syslog 원본을 처리하고 수집할 수 있습니다. 가장 일반적으로 발생하는 두 가지 `rsyslog`, `syslog-ng` 및 Fastvue는 볼륨이 적은 환경과 대규모 분산 환경에서 확장 가능하고 간편하게 통합 및 관리할 수 있는 상용 및 무료 솔루션을 제공합니다.

모니터 설정 방법에 대한 자세한 내용은 데이터 가져오기의 파일 및 디렉터리 모니터링을 참조하십시오.

이 아키텍처는 유니버설 포워더가 다른 엔드포인트에서 수행하는 방식과 동일하게 적절한 데이터 온보딩을 지원합니다. 여러 다양한 로그 유형을 식별하고 적절한 파일 및 디렉터리(Splunk가 전달 작업 시 해당 파일과 디렉터리를 선택할 수 있음)에서 로그 이벤트를 작성하도록 SCD를 설정할 수 있습니다. 또한 디스크에 이벤트를 기록하여 syslog 로그 스트림에 지속성 레벨을 추가하면 신뢰할 수 없는 UDP를 전송 수단으로 사용하여 전송된 메시지의 데이터 손실 위험을 방지할 수 있습니다.

UF를 사용하는 Syslog 데이터 컬렉션 토폴로지



이 다이어그램은 포트 514의 TCP 또는 UDP를 사용하여 syslog 서버의 부하 분산된 풀로 데이터를 전송하는 syslog 원본을 보여줍니다. 여러 서버를 사용하면 컬렉션 계층의 HA를 보장하고 메인터넌스 작업 중 발생하는 데이터 손실을 방지할 수 있습니다. 각 syslog 서버는 syslog 이벤트에 규칙을 적용하여 syslog 이벤트가 각 원본 유형(방화벽 이벤트, OS syslog, 네트워크 스위치, IPS 등)에 대한 전용 파일/디렉터리에 작성되도록 설정됩니다. 각 서버에 배포된 UF는 해당 파일을 모니터링하고 데이터를 인덱싱 계층으로 전달하여 적절한 인덱스로 처리합니다. Splunk AutoLB는 사용 가능한 인덱서 전체에 걸쳐 데이터를 균등하게 분산하는 데 사용됩니다.

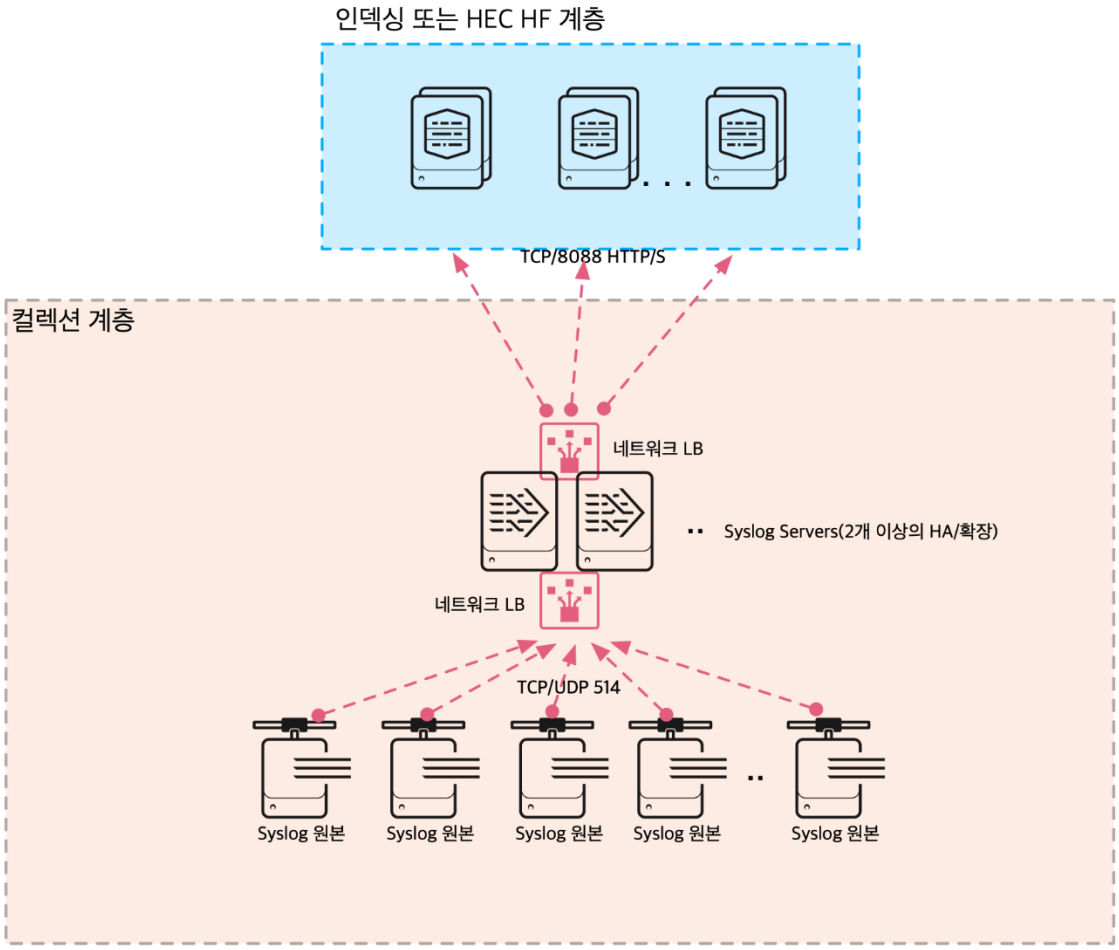
관리 계층에 표시된 배포 서버를 사용하여 UF 설정을 중앙에서 관리할 수 있습니다.

HEC에 대한 Syslog 에이전트

점차적으로 HEC 도입이 늘면서 HEC 배포를 활용하여 syslog를 수집하는 배포 수도 증가하고 있습니다. 자세한 내용은 Splunk 블로그 게시물 Syslog-ng 및 HEC: Splunk의 확장 가능한 집계 데이터 컬렉션을 참조하십시오.

아래 다이어그램은 네트워크 부하 분산 장치를 사용하여 syslog 서버 팜에 포트 514에서 데이터를 보내는 syslog 원본을 보여줍니다. 사용자 지정 syslog 대상(HEC API를 사용하는 Python 스크립트)이 있는 적절한 syslog 정책이 적용되고 이벤트는 HEC 수신기에 전송되며 인덱싱을 위한 네트워크 트래픽 부하 분산기도 함께 전송됩니다.

HEC를 사용하는 Syslog 데이터 컬렉션 토폴로지



이 토폴로지의 장점은 UF/HF를 배포하고 설정할 필요가 없다는 점입니다. HTTP 부하 분산기를 사용하면 인덱서(또는 전용 HEC 수신기 계층)에서 HEC 수신기를 제공하여 데이터를 HEC 엔드포인트에 균등하게 분산할 수 있습니다. "최소 연결" 정책을 사용하여 이 부하 분산기를 설정합니다.

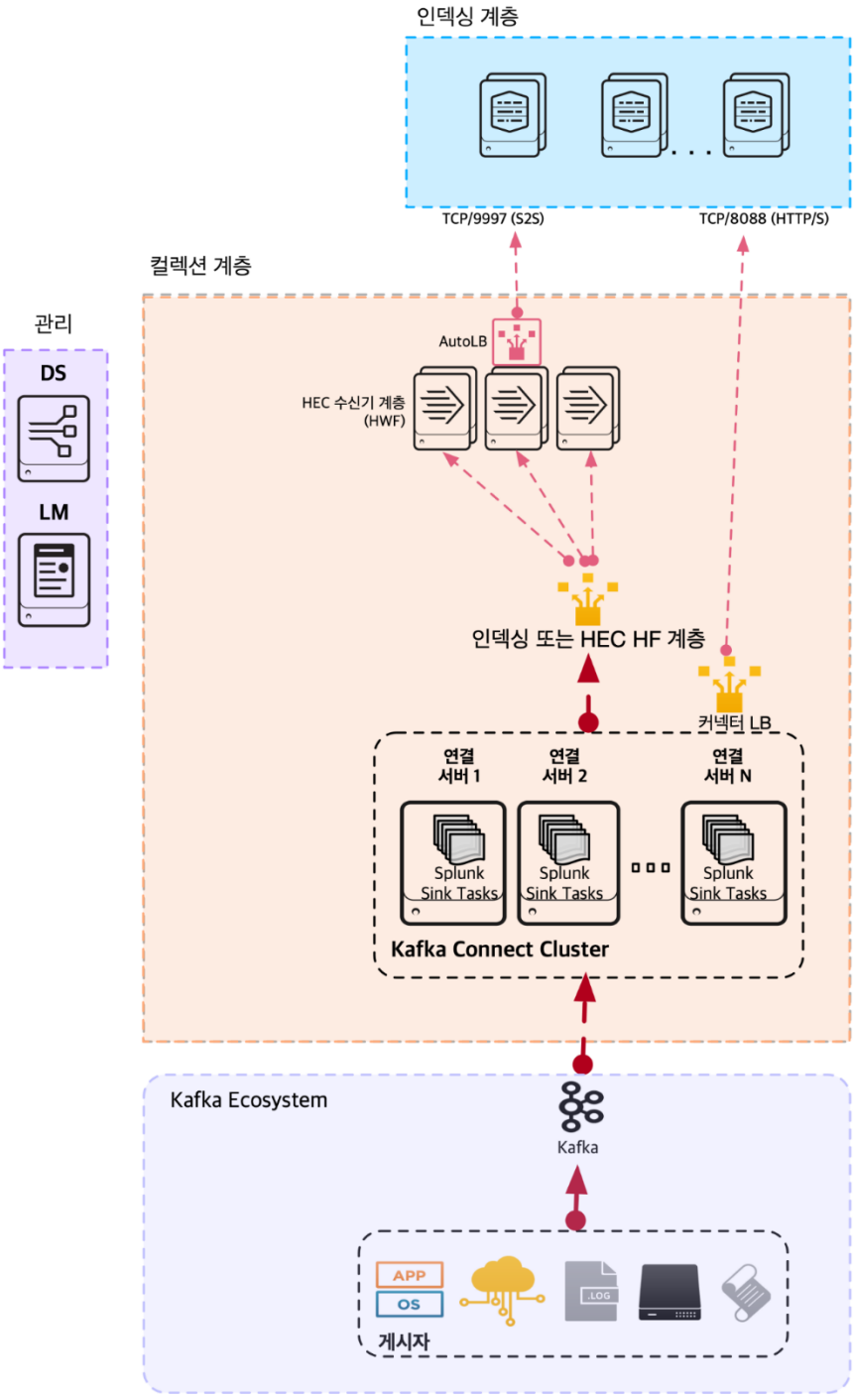
Splunk UDP 입력

Splunk는 UF 또는 HF에서 직접 UDP 입력을 사용하여 syslog에서 데이터를 수신할 수 있습니다. TCP 및 UDP 포트 설정에 대한 자세한 내용은 데이터 가져오기의.TCP 및 UDP 포트에서 데이터 가져오기를 참조하십시오. UDP 514에서 이벤트를 수신하는 기능은 UF/HF가 루트로 실행하는 기능에 의존합니다. 또한 데이터 손실을 방지할 수 있도록 에이전트를 100% 활용할 수 있어야 합니다. 포워더는 설정 변경 사항을 적용하기 위해 빈번하게 다시 시작할 수 있으며, 이를 통해 거의 모든 데이터 손실을 방지할 수 있습니다. 이러한 이유로 **운영 배포의 모범 사례로 간주되지 않습니다.**

Kafka 항목에서 로그 데이터 사용(KAFKA)

Splunk는 "Splunk Connect for Kafka"라는 Kafka 항목에서 데이터를 사용하는 데 지원되는 싱크 커넥터를 제공합니다. 자세한 제품 매뉴얼을 보려면 Splunk Connect for Kafka 매뉴얼의 Apache Kafka Connect를 참조하십시오. Splunk Connect for Kafka 패키지는 적절한 크기의 Kafka Connect 클러스터(Splunk 외부에 위치)에 설치되며, 여기에 설정된 항목을 구독할 수 있으며 HEC를 통해 사용한 이벤트를 전송하여 인덱싱할 수 있습니다.

Kafka 및 HEC를 사용하는 데이터 컬렉션 토폴로지

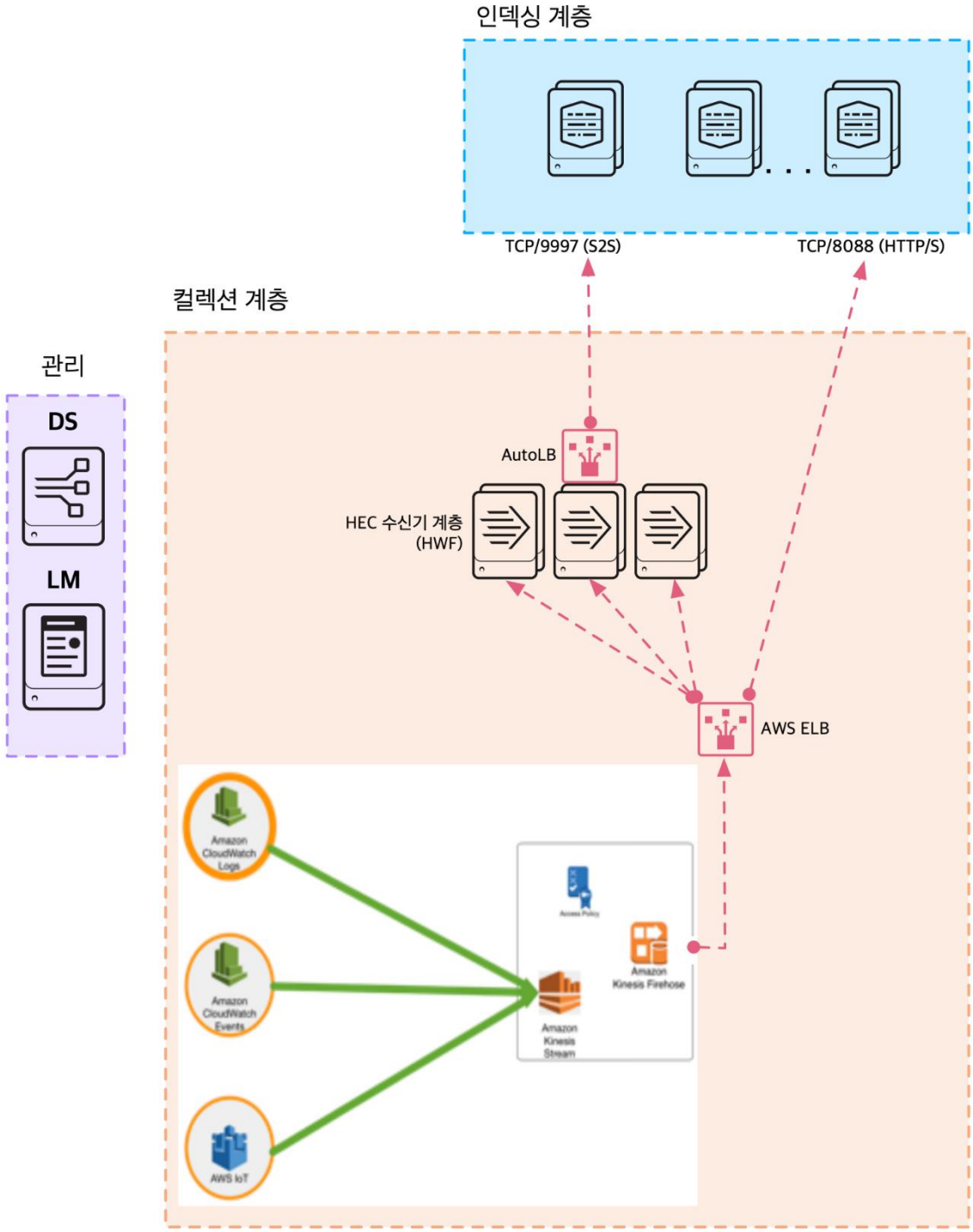


이 다이어그램은 Kafka 게시자가 Kafka 버스로 메시지를 전송함을 보여줍니다. Kafka Connect 클러스터에서 호스트된 작업은 Splunk Connect for Kafka를 통해 해당 메시지를 사용하고 네트워크 부하 분산기를 사용하여 HEC 수신 서비스로 데이터를 전송합니다. 즉, HEC 수신 서비스는 인덱서에서 직접 호스트되거나 전용 HEC 수신기 계층에서 호스트될 수 있습니다. 자세한 내용은 HEC 절을 참조하십시오. 관리 계층 구성 요소는 전용 HF 계층이 호스트 HEC 수신기에 배포되는 경우에만 필요합니다.

Amazon Kinesis Firehose의 로그 데이터 사용(KINESIS)

Splunk와 Amazon은 Kinesis와 Splunk HEC 간의 통합을 구현하여 AWS에서 AWS 콘솔을 통해 설정 가능한 HEC 엔드포인트로 데이터를 직접 스트리밍할 수 있습니다. 이렇게 하면 AWS에서 생성된 다양한 데이터 원본에 대한 CIM 호환 지식을 제공하는 Splunk Add-On for Kinesis Firehose로 보완됩니다.

Amazon Kinesis를 사용하는 데이터 컬렉션 토폴로지



이 다이어그램은 Kinesis 스트림을 사용하여 Firehose에 전송되는 AWS 로그 원본을 보여주며, 적절하게 설정된 Firehose는 AWS ELB를 통해 HEC 수신 서비스로 데이터를 전송합니다. 즉, HEC 수신 서비스는 인덱서에서 직접 호스트되거나 전용 HEC 수신기 계층에서 호스트될 수 있습니다. 자세한 내용은 HEC 절을 참조하십시오.

표시된 관리 계층 구성 요소는 전용 HF 계층이 호스트 HEC 수신기에 배포된 경우에만 필요합니다.

메트릭 컬렉션(METRICS)

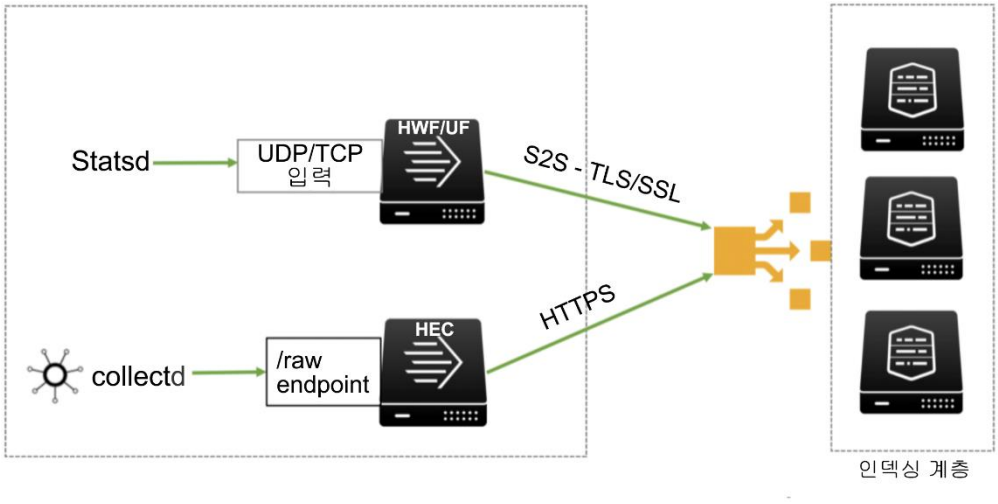
Splunk는 다양한 타사 소프트웨어에서 시스템 및 애플리케이션 성능 데이터 또는 메트릭 데이터를 수신 및 수집할 수 있습니다. Splunk 플랫폼의 메트릭은 메트릭 저장 및 검색에 최적화된 사용자 지정 인덱스 유형을 사용합니다.

메트릭 데이터를 사용하는 다양한 방법이 있으며 컬렉션 방법은 사용된 기술을 기반으로 합니다. 가장 일반적인 메트릭 컬렉션 양식은 **collectd**, **statsd** 같은 소프트웨어 데몬 형식으로 제공되거나 사용자 지정 메트릭 데이터 파일과 데이터 원본에 대한 유효한 설정을 사용합니다.

Splunk로 메트릭을 수집하는 두 가지 주요 방법에는 **statsd**와 **collectd** 같은 에이전트를 사용하는 방법이 있습니다. **직접 TCP/UDP** 입력을 사용하거나 **HEC**를 통해 수집합니다.

컬렉션 계층을 수평적으로 간편하게 확장할 수 있는 기능과 **HEC** 엔드포인트의 복원력과 확장성을 고려하면 **HEC**를 사용하는 것이 가장 좋습니다.

메트릭 데이터 컬렉션 토폴로지



Statsd는 현재 **UDP**와 **TCP** 전송 방식을 지원하며, 이 전송 방식은 Splunk 포워더 또는 인덱서에서 직접 입력으로 사용할 수 있습니다. 하지만 아키텍처가 복원되지 않고, 필수 Splunk 포워더의 다시 시작으로 인해 이벤트 손실(Syslog 컬렉션 참조)이 발생할 수 있으므로 TCP/UDP 트래픽을 운영 중인 포워더에 직접 전송하는 것이 좋습니다.

계층 구성 요소 전달을 위한 고가용성(HA) 고려 사항

디지털 환경에는 고가용성(HA)이라는 공통된 개념이 있습니다. 하지만 조직에 따라 의미는 달라질 수 있으며, 이 의미는 고가용성과 달리 재해 복구(DR)에 더 부합합니다. 이 두 개념은 유사하지만 다른 의미를 가지고 있습니다. HA는 평상 시보다 높은 합의된 수준의 운영 성능(일반적으로 가동 시간)을 보장하는 것을 목표로 하는 시스템 특성입니다. DR에는 재난 발생 후 핵심 기술 인프라와 시스템을 복구하거나 지속할 수 있는 일련의 정책, 도구 및 절차가 포함됩니다.

다음은 중간/집계 계층에서 다양한 형태의 HA를 설명합니다.

중간 계층

- 중간/집계 계층 배포를 사용하는 고객에게 포워더의 HA는 필수적입니다. 애플리케이션 계층에서 Splunk에서 현재 HA에 대해 기본적으로 지원하는 방법은 없으며, Splunk 고유가 아닌 운영 체제 레벨에서 HA를 제공하는 다른 전략이 있습니다. 일반적인 솔루션에는 VMware VMotion, AWS Autoscaling Groups 및 Linux Clustering이 포함됩니다. Splunk Architect에 문의하여 사용 가능한 다른 설계 옵션에 대해 논의하십시오.
- 전용 HEC 계층에 대한 HA 요구 사항이 있는 환경의 경우 Splunk의 여러 헤비 포워더 앞에 NGINX 같은 네트워크 트래픽 부하 분산기(NTLB)를 사용하는 것이 가장 좋습니다. 이는 최대 처리량, 규모 및 가용성의 이점을 제공합니다. 데이터를 수신하고 전달하는 작업만 수행하는 전용 HTTP Event Collector 인스턴스 풀이 있습니다. 인덱서를 더 추가할 필요 없이 더 많은 HEC 인스턴스를 추가할 수 있습니다. 인덱서에 병목 현상이 발생하면 추가 인덱서를 추가합니다.
- Syslog 컬렉션에 대한 HA 요구 사항이 있는 환경의 경우 HAProxy 또는 F5 같은 부하 분산 솔루션이 호스트된 클러스터(가상) IP 주소로 제공되는 여러 Syslog 서버를 사용하여 최대 처리량, 규모 및 가용성의 이점을 제공하는 것이 가장 좋습니다. 데이터를 수신하고 전달하는 작업만 수행하는 전용 Splunk 인스턴스 풀이 있습니다. 인덱서를 더 추가할 필요없이 인스턴스를 추가할 수 있습니다. 인덱서에 병목 현상이 발생하면 추가 인덱서를 추가합니다.

전달 계층

- 전달(엔드포인트) 계층에서 에이전트 자체에 대한 HA는 기본 OS에 종속됩니다. 최소한 호스트 운영 체제가 다시 시작될 때 전달 기능을 구현하는 모든 서비스가 자동으로 다시 시작되어야 합니다. 이 외에도 포워더의 경우 포워더에서 여러 인덱서에 이르기까지 AutoLB를 적절히 사용하고 설정하는 것이 좋습니다. 또한 인덱싱 계층에 데이터가 도달하도록 인덱서 수신확인 기능을 사용하는 것이 좋습니다.

3단계: 설계 원리 및 모범 사례 적용

아래에서는 배포 계층에 따라 설계 원리와 모범 사례를 구분합니다.

배포 계층

SVA 설계 원리는 다음과 같이 모든 배포 계층을 포함합니다.

계층	정의
검색	• 검색 헤드
인덱싱	• 인덱서
컬렉션	• 포워더

	• 모듈식 입력 • 네트워크 • HTTP Event Collector(HEC) • 등
관리/유틸리티	• CM • DS • LM • DMS • SHC-D

토폴로지와 모범 사례 연결

배포에 적합한 설계 원리와 모범 사례를 선택하려면 요구 사항과 토폴로지를 고려해야 합니다. 따라서 위의 Splunk Validated Architecture 선별 프로세스의 1단계와 2단계를 완료한 후에 모범 사례만 고려해야 합니다.

모범 사례: 계층별 권장 사항

아래에서 각 배포 계층에 따라 설계 원리와 모범 사례 권장 사항을 확인할 수 있습니다. 각 설계 원리는 가용성, 성능, 확장성, 보안 및 관리 기능 같은 하나 이상의 SVA 기본 요소를 보완합니다.

검색 계층 권장 사항

설계 원리/모범 사례 (사용자의 요구 사항에 따라 적용되는 모범 사례가 결정됩니다)		**SVA 기본 요소**				
		가용성	**성능**	**확장성**	**보안**	**관리 기능**
1	검색 계층을 인덱싱 계층에 가까운 위치에 배치(네트워크 측면에서) 검색 계층과 인덱싱 계층 간의 모든 네트워크 지연은 검색 성능에 직접적인 영향을 미칩니다.		✓			
2	여러 개의 독립적인 검색 헤드 사용 지양 독립적인 검색 헤드는 사용자가 만든 Splunk 아티팩트를 공유할 수 없습니다. 또한 전체 검색 계층에 걸쳐 리소스 활용과 관련하여	✓		✓	✓	✓

	적절히 확장되지 않습니다. 분리된 검색 헤드 환경을 특별히 필요로 하지 않는 한 확장하기 위한 더 나은 옵션이 있습니다.					
3	검색 계층을 확장할 때 검색 헤드 클러스터링 사용 검색 헤드 클러스터는 전체 클러스터에서 사용자 아티팩트를 복제하고 클러스터의 모든 구성원에게 지능형 검색 워크로드 스케줄링을 허용하며, 고가용성 솔루션도 제공합니다.	✅		✅		
4	모든 검색 헤드의 내부 로그를 인덱싱 계층으로 전달 인덱싱된 모든 데이터는 인덱싱 계층에만 저장해야 합니다. 이를 통해 검색 헤드 계층에 고성능 스토리지를 제공할 필요가 없으며 관리가 용이해집니다. 참고: 다른 Splunk 역할에도 적용됩니다.		✅			✅
5	가능한 경우 LDAP 인증 사용 고려 인증하려면 사용자 ID를 중앙에서 관리하는 것은 일반적인 엔터프라이즈 모범 사례이며 Splunk 배포 관리를 단순화하고 보안을 강화합니다.				✅	✅
6	동시 검색 요구 사항을 충족할 만큼 충분한 코어 확보 모든 검색을 수행하려면 CPU 코어가 필요합니다. 검색을 실행할 수 있는 코어가 없으면 검색은 대기 상태가 되어 사용자에 대한 검색 지연이 발생할 수 있습니다. 참고: 인덱싱 계층에도 적용됩니다.	✅	✅	✅		

7	가능한 한 예약된 검색 시간 윈도우 사용/원활한 예약된 검색 부하 종종 예약된 검색은 특정 시간대(정시, 5/15/30분 후, 자정)에 실행됩니다. 검색을 실행할 수 있는 시간 윈도우를 제공하면 검색 동시성 핫스팟을 방지할 수 있습니다.		✓	✓		
9	인덱스 계층을 초과하지 않도록 개별 검색 헤드 클러스터 수 제한 검색 워크로드는 SH 환경 내에서만 자동으로 관리됩니다. 독립적인 SHC는 인덱서(검색 피어) 계층이 처리할 수 있는 워크로드보다 더 많은 동시 검색 워크로드를 만들 수 있습니다. 독립형 검색 헤드의 수를 신중하게 계획하는 경우에도 마찬가지입니다.	✓		✓		
10	검색 헤드 클러스터를 만들 경우 홀수 노드(3, 5, 7 등) 사용 SHC 캡틴 선출은 다수에 기반한 프로토콜을 통해 수행됩니다. 노드 수가 홀수일 경우 네트워크 오류 발생 시 SHC를 짝수 개의 노드로 분할할 수 없습니다.	✓				✓

인덱싱 계층 권장 사항

설계 원리/모범 사례 (사용자의 요구 사항에 따라 적용되는 모범 사례가 결정됩니다)		**기본 요소**				
		가용성	**성능**	**확장성**	**보안**	**관리 기능**
1	사용 가능한 서버에서 병렬식 파이프라인 활성화 병렬식 처리 기능을 사용하면 유휴 상태에 놓일 수 있는 사용 가능한		✓	✓		

	시스템 리소스를 악용할 수 있습니다. 수집 병렬 처리 기능을 사용하려면 I/O 성능이 양호해야 합니다.					
2	HOT/WARM 볼륨 및 요약에 SSD 사용 고려 SSD는 경제적인 가격으로 제공되며 불만족스러운 검색 성능을 야기할 수 있는 IO 제한 요소를 제거합니다.		✓			
3	인덱싱 계층을 검색 계층에 가까운 위치에 배치(네트워크 측면에서) 가능한 한 가장 짧은 네트워크 대기 시간은 검색 시 사용자 환경에 긍정적인 영향을 미칩니다.		✓			
4	내역 데이터/보고서 HA가 필요한 경우 인덱스 복제 사용 인덱스 복제는 검색 피어 오류가 발생하지 않도록 클러스터의 모든 이벤트 복사본을 여러 개 만듭니다. SLA에 맞게 복사본(복제 팩터) 수를 조정합니다.	✓				
5	양호한 데이터 온보딩 상태(예: 줄 바꿈, 타임스탬프 추출, TZ 및 원본, 원본 유형, 호스트가 각 데이터 원본에 대해 적절하게 명시적으로 정의됨)를 확인하고 모니터링 콘솔을 사용하여 지속적인 모니터링 설정 데이터 원본을 명시적으로 설정하지 않고 Splunk의 자동 탐지 기능에 의존하는 것은 데이터 수집 용량 및 인덱싱 대기 시간, 특히 대용량 배포의 경우 상당한 이점이 있는 것으로 입증되었습니다.		✓	✓		✓
6	초과 처리 기능으로 인덱서에서 배치 모드 검색 병렬 처리 설정을 설정하도록 고려		✓	✓		

	검색 병렬 처리 기능을 활용하면 특정 유형의 검색에 대한 검색 성능에 상당한 영향을 미칠 수 있으며, 그렇지 않은 경우 사용하지 않을 수도 있는 시스템 리소스를 활용할 수 있습니다.					
7	인덱서 노드(= 검색 피어)에 걸쳐 분산된 데이터 분산 모니터링 검색 피어의 이벤트/데이터 분산조차 검색 성능 및 적절한 데이터 보존 정책 실행에 필요한 중요한 기여 요인입니다.		✅	✅		✅
8	분산된/클러스터된 배포의 인덱서에서 웹 UI 비활성화 인덱서에서 WebUI에 직접 액세스할 필요는 없습니다.		✅		✅	✅
9	잘 알려진 데이터 원본에 대해 Splunk에서 기본으로 제공되는 기술 추가 기능 고려 잘 알려진 데이터 원본에 대해 데이터 온보딩 상태를 보장하려면 자체 설정을 만드는 대신, Splunk 제공 TA를 사용하면 가치 실현 시간을 단축하고 최적의 구현을 보장할 수 있습니다.		✅			✅
10	중요한 인덱서 메트릭 모니터링 Splunk는 인덱싱 계층의 수행 방식에 대한 주요 성능 메트릭을 제공하는 모니터링 콘솔을 지원합니다. 여기에는 내부 Splunk 구성 요소(프로세스, 파이프라인, 대기열, 검색)의 상세 메트릭뿐 아니라 CPU 및 메모리 활용도가 포함됩니다.	✅	✅			

컬렉션 계층 권장 사항

	설계 원리/모범 사례 (사용자의 요구 사항에 따라 적용되는 모범 사례가 결정됩니다)	기본 요소				
		가용성	성능	확장성	보안	관리 기능
1	가능하면 UF를 사용하여 데이터를 전달합니다. 헤비 포워더의 사용 방법은 이 포워더를 필요로 하는 이용 사례로 제한되어야 합니다. 기본 제공 autoLB, 다시 시작 가능, 중앙에서 설정 가능, 소규모 리소스 요구		✓			✓
2	많은 UF를 유입할 때 인덱서에 최소 2x개 이상의 중간 전달 파이프라인 사용 적은 수의 중간 포워더에 다수의 엔드포인트 포워더를 멀티플렉싱하면 인덱서에서 이벤트 분산에도 영향을 미치며, 이는 검색 성능에도 영향을 미치게 됩니다. **반드시 필요한 경우 중간 포워더만 배포합니다.**	✓	✓			
3	SSL을 사용하여 UF-IDX 트래픽 보안 고려				✓	
4	기본 Splunk LB를 사용하여 데이터를 인덱싱 계층에 분산 네트워크 부하 분산기는 현재 포워더와 인덱서 간에 지원되지 않습니다.	✓		✓		
5	Syslog 컬렉션 전용 syslog 서버 사용 Syslog 서버는 원본을 기반으로 디스크에 TCP/UDP 트래픽을 유지하고 유니버설 포워더를 사용하여 적절한 원본 유형 설정을 처리할 수 있습니다.	✓				✓

	요청된 포워더 다시 시작으로 인해 데이터가 손실되지 않습니다.					
6	에이전트가 없는 컬렉션에 원시 TCP/UDP 대신 HEC 사용 HTTP Event Collector(HEC)는 HTTP[/S] 프로토콜을 통해 이벤트를 게시할 수 있는 수신 서비스입니다. 인덱서에서 직접 활성화하거나 헤비 포워더 계층에서 설정할 수 있으며, 모두 부하 분산기에서 제공됩니다.	✓				✓

관리/유틸리티 계층 권장 사항

설계 원리/모범 사례 (사용자의 요구 사항에 따라 적용되는 모범 사례가 결정됩니다)		**기본 요소**				
		가용성	**성능**	**확장성**	**보안**	**관리 기능**
1	소규모 환경을 위한 단일 인스턴스에서 LM, CM, SHC-D 및 MC 통합 고려 이러한 서버 역할은 리소스 요구량이 거의 없고 코로케이션에 적합한 후보입니다. 대규모 인덱서 클러스터에서 CM은 클러스터를 효율적으로 관리하기 위한 전용 서버가 필요할 수 있습니다.					✓
2	중간 규모 이상의 대규모 배포를 위해 DS에 대한 별도의 인스턴스 고려 상당 수의 포워더가 배포 서버를 통해 관리되면 리소스 요구량은 전용 서버가 서비스를 유지 관리하는 데 필요한 위치로 증가합니다.					✓
3	대규모 배포의 경우 LB 이후 여러 DS 고려	✓		✓		✓

	참고: 올바르게 설치 및 설정하려면 Splunk 프로페셔널 서비스의 지원을 받아야 할 수도 있습니다.					
4	DS phoneHomeIntervalInSecs를 60초 기본값에서 백업할 수 있을지 여부 결정 전화 홈 간격이 길면 DS 확장성에 긍정적인 영향을 미칩니다.			✓		
5	전용/보안 DS를 사용하여 앱 배포를 통한 클라이언트 악용 방지 배포 서버에 액세스할 수 있는 모든 사용자는 DS에 의해 관리되는 Splunk 설정을 수정할 수 있습니다. 여기에는 잠재적으로 포워더 엔드포인트에 악성 애플리케이션을 배포하는 것도 포함됩니다. 이 역할을 안전하고 적절하게 유지하는 것이 중요합니다.				✓	
6	모니터링 콘솔(MC)을 사용하여 배포 상태를 모니터링하고 상태 관련 문제에 대해 경고 전송 모니터링 콘솔은 사전에 작성한 Splunk의 모니터링 솔루션 집합을 제공하며 사용자 환경의 성능을 저해하는 요소에 대해 사용자에게 알림을 보내는 확장 가능한 플랫폼 경고를 포함합니다.	✓	✓			✓

요약 및 다음 단계

이 백서는 Splunk Validated Architecture에 대한 일반적인 내용을 소개합니다. Validated Architecture는 가장 경제적이고 관리 및 확장 가능한 방식으로 조직의 요구 사항이 충족되도록 보장합니다. SVA는 다음과 같은 기본 요소에 기반한 모범 사례와 설계 원리를 제공합니다.

- 가용성
- 성능
- 확장성
- 보안
- 관리 기능

이 백서에서는 다음과 같이 3단계의 Splunk Validated Architecture 선별 프로세스를 다룹니다. 3단계는 (1) 요구 사항 정의, 2) 토폴로지 선택, 3) 설계 원리 및 모범 사례 적용입니다. Splunk Validated Architecture의 다양한 이점을 알아보았습니다. 이제 조직에 적합한 배포 토폴로지를 선택하는 프로세스를 진행할 준비를 마쳤습니다.

다음 단계

Validated Architecture를 선택한 후에는 어떤 절차를 수행해야 합니까? 작업 환경에 대해 수행할 이후 절차는 다음과 같습니다.

사용자 지정

- 선택한 토폴로지가 특정 요구 사항을 충족시키는 데 필요할 수 있는 필수 사용자 지정 항목을 고려합니다.

배포 모델

- 배포 모델(베어메탈, 가상, 클라우드)을 결정합니다.

시스템

- Splunk 시스템 요구 사항에 따라 기술(서버, 스토리지, 운영 체제)을 선택합니다.

규모 지정

- 배포 규모를 지정하는 데 필요한 관련 데이터(데이터 수집, 예상 검색량, 데이터 보존 요구, 복제 등)를 수집합니다. Splunk 스토리지 규모 지정(https://splunk-sizing.appspot.com/)은 사용 가능한 도구 중 하나입니다.

인력 충원

- 배포를 구현 및 관리하기 위한 인력 수요를 평가합니다. 이는 Splunk Center of Excellence를 구축하는 데 필수적인 부분입니다.

Splunk는 Validated Architecture 프로세스와 다음 단계를 통해 고객을 지원합니다. 문의 사항이 있는 경우 Splunk Account Team에 문의하시기 바랍니다. 사용자의 계정 팀은 Splunk 내 모든 기술 및 아키텍처 리소스에 액세스할 수 있으며 상세 정보를 제공할 수 있습니다.

Splunk와 함께 하세요!

부록

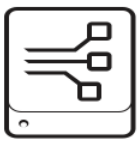
이 절에서는 SVA에서 사용되는 추가 참조 정보를 확인할 수 있습니다.

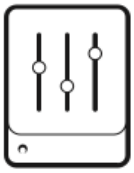
부록 "A": SVA 기본 요소 설명

기본 요소	설명	기본 목표/설계 원리
가용성	지속적으로 운영되거나 계획된 작동 중단 또는 예상치 못한 작동 중단으로부터 복구할 수 있는 능력입니다.	1. 단일 장애 발생 지점 제거/중복 추가 2. 계획된 장애/작동 중단 및 예상치 못한 장애/작동 중단 탐지 3. 계획된/예상치 못한 작동 중단을 이상적으로 자동 허용 4. 롤링 업그레이드 계획
성능	다양한 사용 패턴에서 최적의 서비스 레벨을 유지하기 위해 가용 리소스를 효과적으로 활용할 수 있는 기능입니다.	1. 성능 향상을 위해 하드웨어를 추가합니다(계산, 스토리지, 메모리). 2. '아주 처음 단계'부터 시작된 병목 현상 제거 3. 동시 처리 수단을 모두 사용 4. 주변 요소(예: 구성 요소의 분산 최소화) 활용 5. 일반적인 사례에 맞게 최적화(80/20 규칙) 6. 불필요한 일반성을 피함 7. 시간 이동 계산(사전 계산, 지연 계산, 공유/배치 계산) 8. 시간의 확실성과 정확성 교차(무작위 추출, 샘플링)
확장성	시스템이 모든 계층에서 확장되고 증가된 워크로드를 효율적으로 처리하도록 설계된 기능입니다.	1. 세로 및 가로 비율 조정 2. 개별적으로 확장해야 하는 개별 기능 구성 요소 3. 구성 요소 간 종속성 최소화

기본 요소	설명	기본 목표/설계 원리
		4. 가능한 한 신속하게 미래 성장을 위한 설계 수행 5. 전반적인 시스템 설계의 계층 구조 소개
보안	시스템이 지속적으로 가치를 제공하고 설정 및 자산뿐 아니라 데이터를 보호하도록 설계된 기능입니다.	1. 시작부터 안전한 시스템을 위한 설계 2. 모든 통신에 최첨단 프로토콜 사용 3. 이벤트 데이터에 대한 광범위한 수준의 세분화된 액세스 허용 4. 중앙 집중식 인증 사용 5. 감사 절차 구현 6. 공격이나 악의적인 사용 줄이기
관리 기능	시스템을 보장하는 기능은 모든 계층에서 중앙에서 운영 및 관리할 수 있도록 설계되었습니다.	1. 중앙 집중식 관리 기능 제공 2. 설정 개체 라이프사이클 관리(원본 제어) 3. 애플리케이션(Splunk) 사용 측정 및 모니터링/프로파일링 4. 시스템 상태 측정 및 모니터링

부록 "B": 토폴로지 구성 요소

계층	구성 요소	아이콘	설명	참고
관리	배포 서버(DS)		배포 서버는 포워더 설정의 설정을 관리하며,	전용 인스턴스에 배포해야 합니다. 또한 간편한 장애 복구를 위해 가상화할 수 있습니다.
	라이선스 마스터(LM)		라이선스 마스터는 기타 Splunk 구성 요소가 라이선스 기능을 사용하고 일별 데이터 수집 볼륨을 추적하는 데 필요합니다.	라이선스 마스터의 역할은 최소한의 용량 및 가용성 요구 사항을 가지며 기타 관리 기능과 코로케이션될 수 있습니다. 또한 간편한

계층	구성 요소	아이콘	설명	참고
				장애 복구를 위해 가상화할 수 있습니다.
	모니터링 콘솔(MC)		모니터링 콘솔은 사용자 환경의 사용 및 상태 모니터링을 위한 대시보드를 제공합니다. 또한 작동 문제에 대한 알림을 제공하도록 사용자 지정할 수 있는 사전에 패키징된 수많은 플랫폼 경고를 포함합니다.	클러스터된 환경에서 MC는 클러스터되지 않은 배포의 라이선스 마스터 및 배포 서버 기능 외에도 마스터 노드와 코로케이션할 수 있습니다. 또한 간편한 장애 복구를 위해 가상화할 수 있습니다.
	클러스터 마스터(CM)		클러스터 마스터는 클러스터된 배포의 모든 작업에 필요한 코디네이터입니다.	다수의 인덱스 버킷(높은 데이터 볼륨/보존)이 있는 클러스터에서 클러스터 마스터는 전용 서버를 실행해야 할 수 있습니다. 또한 간편한 장애 복구를 위해 가상화할 수 있습니다.
	검색 헤드 클러스터 주 배포 노드(SHC-D)		검색 헤드 클러스터 주 배포 노드는 SHC를 부트스트랩하고 클러스터에 배포된 Splunk 설정을 관리하는 데 필요합니다.	SHC-D는 런타임 구성 요소가 아니며 시스템 요구 사항이 거의 없습니다. 기타 관리 역할과 코로케이션할 수 있습니다. 참고: 각 SHC는 자체의 SHC 주 배포 노드 기능이 필요합니다. 또한 간편한 장애 복구를 위해 가상화할 수 있습니다.
검색	검색 헤드(SH)		검색 헤드는 Splunk 사용자를 위한 UI를 제공하고 예약된 검색 작업을 조정합니다.	검색 헤드는 분산 배포의 전용 Splunk 인스턴스입니다. 검색 헤드는 적절한 CPU 및 메모리 리소스를 사용하여 배포한 경우 쉽게 복구할 수 있도록 가상화할 수 있습니다.

계층	구성 요소	아이콘	설명	참고
	검색 헤드 클러스터(SHC)		검색 헤드 클러스터는 세 개 이상의 클러스터된 검색 헤드의 풀입니다. 검색 헤드 계층에 대한 수평식 확장성 및 작동 중단 시 투명한 사용자 페일오버를 제공합니다.	검색 헤드 클러스터에는 이상적으로 동일한 시스템 사양의 전용 서버가 필요합니다. 검색 헤드 클러스터 구성원은 적절한 CPU 및 메모리 리소스로 배포된 경우 간편하게 장애 복구를 수행하도록 가상화할 수 있습니다.
인덱싱	인덱서		인덱서는 Splunk를 대표하는 속성입니다. 수신 데이터를 처리하고 인덱싱하며 검색 계층에서 시작된 검색 요청을 수행하는 검색 피어로도 사용됩니다.	인덱서는 항상 분산 또는 클러스터된 배포의 전용 서버에 있어야 합니다. 단일 서버 배포에서 인덱서는 검색 UI 및 라이선스 마스터 기능도 제공합니다. 인덱서는 적절한 리소스를 보장할 수 있다면 베어메탈 서버 또는 전용 고성능 가상 시스템에서 가장 높은 성능을 보입니다.
데이터 컬렉션	포워더 및 기타 데이터 컬렉션 구성 요소		데이터 컬렉션과 관련된 모든 구성 요소의 일반 아이콘	여기에는 유니버설 포워더 및 헤비 포워더, 네트워크 데이터 입력 및 기타 데이터 컬렉션 형식(HEC, Kafka 등)이 포함됩니다.