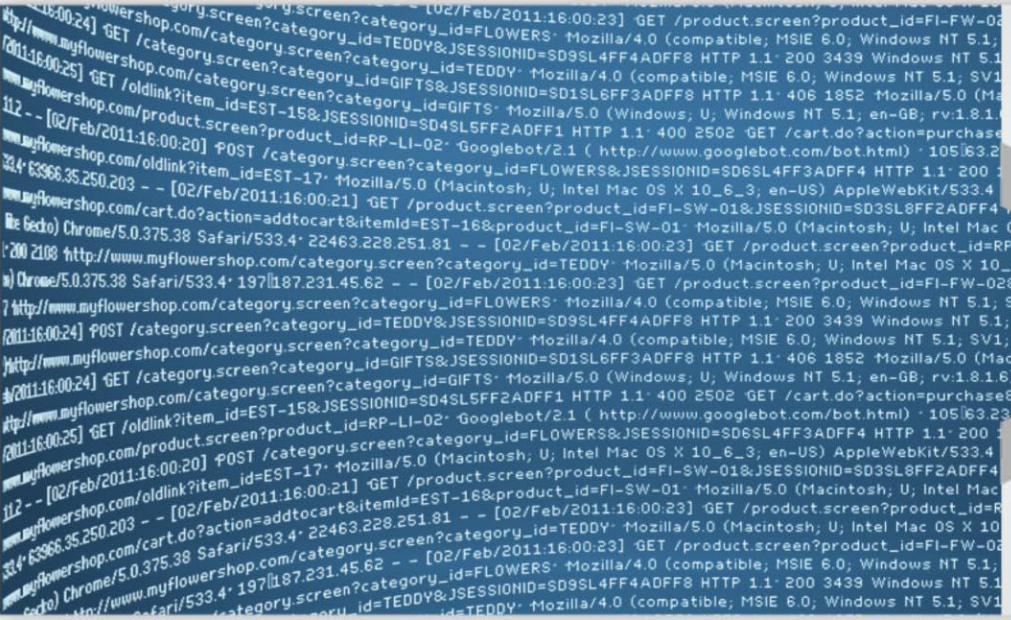
splunk> MOS



Splunk 기능 및 용어 설명

splunk>

Splunk 기능 설명

> Splunk Enterprise

1타 4피!!!

Search header

> 최종 단말 성격의 구성요소
> 검색된 결과 화면을 통합하여 사용자에게 제공

Indexer

> 검색을 실제 수행하는 구성요소
> 인덱스를 생성, 관리 및 데이터 인덱싱 (데이터처리)
> 원시 데이터와 인덱싱된 데이터를 저장 및 검색

Deployment

> 로컬 및 분산 인스턴스 관리

Forwarder

> 데이터를 수집하는 구성요소

Splunk Search Header

Search Header

- **역할,** 검색 요청 관리 및 결과 제공

- **Search Head** 는 여러 Indexers에 검색을 명령하고 결과를 사용자에게 리턴
- **Splunk Daemon** 에 접근하여 검색 수행
- **분산 검색** 시 하나 이상의 검색 헤더가 필요
- **CPU, Memory Resource** 를 상대적으로 많이 사용

Splunk Indexer 관련 용어 설명

Indexer

- **역할,** 데이터 저장 및 검색작업 수행

- **Index** : 데이터 저장소 (RDBMS와 같은 저장소)
  - **main** : 인덱스가 별도로 지정하지 않을 시 Main Index 에 저장됨
  - **_internal** : Splunk Enterprise 내부 로그 및 메트릭이 포함됨
  - **_audit** : 파일 시스템 변경, 감사(audit) 및 모든 사용자 검색 내역과 관련된 이벤트가 포함됨

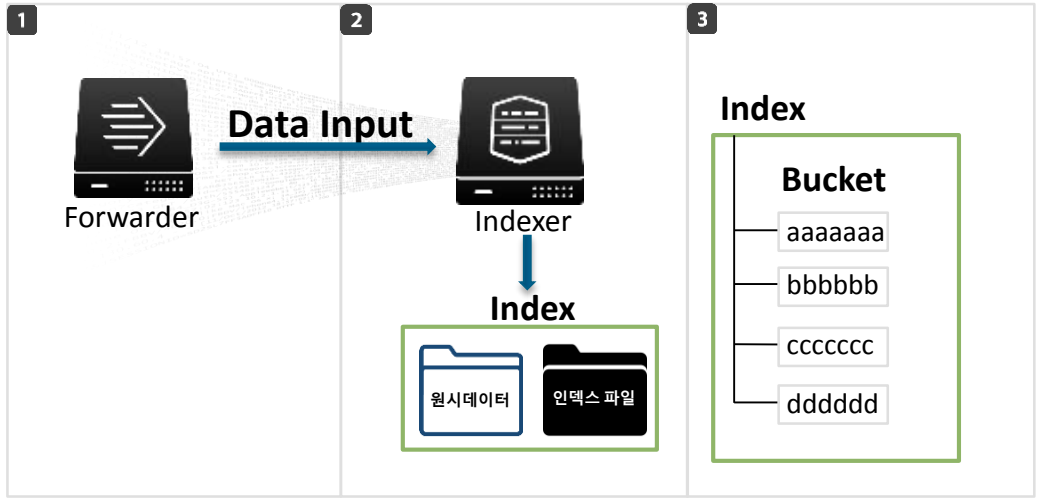
- **Indexing** : 원본 데이터를 Splunk가 검색 가능한 이벤트로 변환하여 저장 기록하는 과정

- **Parsing** : 이벤트를 정규화된 필드로 분류하는 작업 (타임 스템프 식별, regex 규칙에 의한 이벤트 데이터와 메타 데이터 변환 등)

- **Field** : 이벤트를 자동 혹은 정규식으로 Parsing 한 Values의 Key

Splunk Indexer 데이터 저장 구조

> Splunk 데이터 저장



* 인덱스 디렉터리 및 파일의 컬렉션 / 위치 : $SPLUNK_HOME/var/lib/splunk
* 원시 데이터 압축 형 원시 데이터원시 데이터
* 인덱스 파일 원시 데이터, 일부 메타데이터 파일을 가리키는 인덱스
* 버킷 인덱스 디렉터리, 에이지(age)를 기준으로 구성

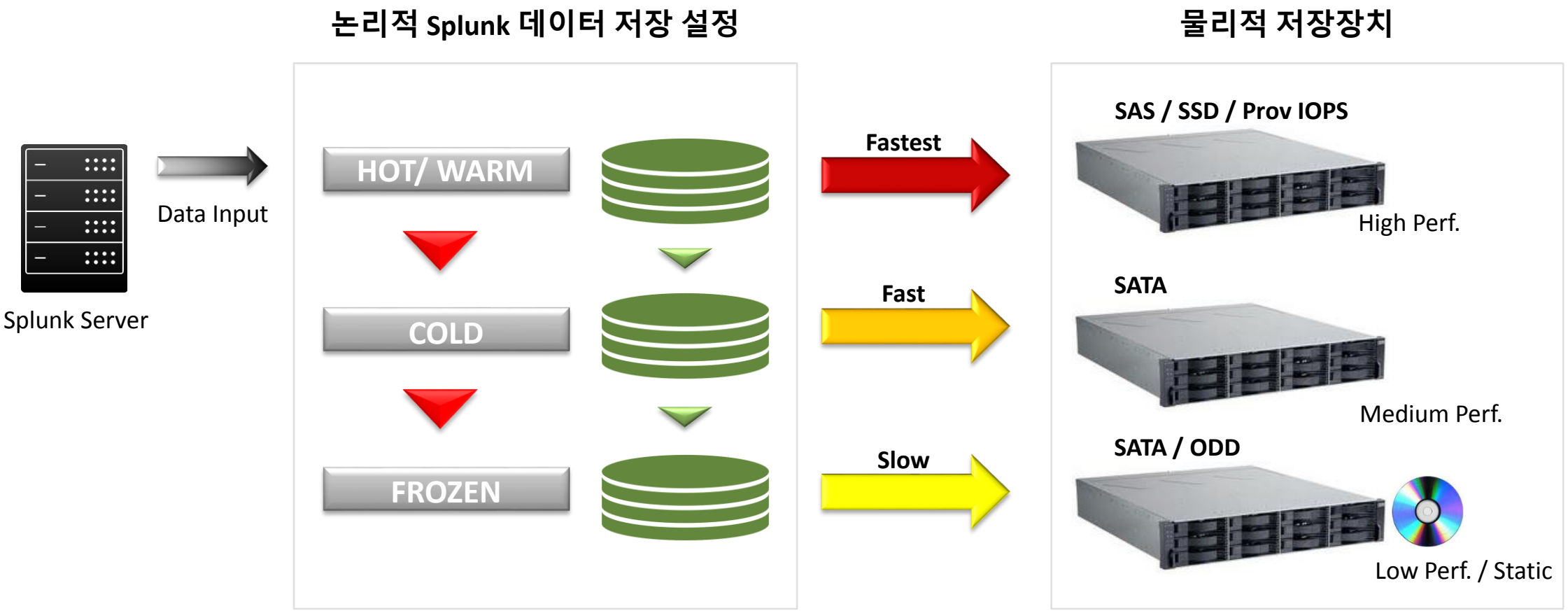
Bucket 단계	event	searchable	설명
Hot	New	yes	새로운 인덱스 데이터가 포함/ 쓰기 허용 각 인덱스에는 하나 이상의 hot 버킷이
Warm	Old	Yes	hot에서 롤링된 데이터입니다.
Cold	Even older	Yes	warm에서 롤링된 데이터입니다.
Frozen	Oldest (delete)	No	cold에서 롤링된 데이터 인덱서는 기본적으로 frozen 데이터를 삭제하지만 아카이브할 수도 있습니다. 아카이브된 데이터를 나중에 해제(thawed)할 수 있습니다.
Thawed	Same as frozen	yes	

Types (and flow) : hot -> warm -> cold -> frozen -> thawed

* Hot : most recent data, small tsidx files, auto splunk-optimize, searchable
* Warm : older than hot, searchable, bloom filter
* Cold : older than warm, searchable, bloom filter (or replicated data, not necessary searchable)
* Frozen : not searchable, deleted by default (can set "coldToFrozenScript" and/or "coldToFrozenDir" to archive the data)
* Thawed : from frozen (archived), searchable

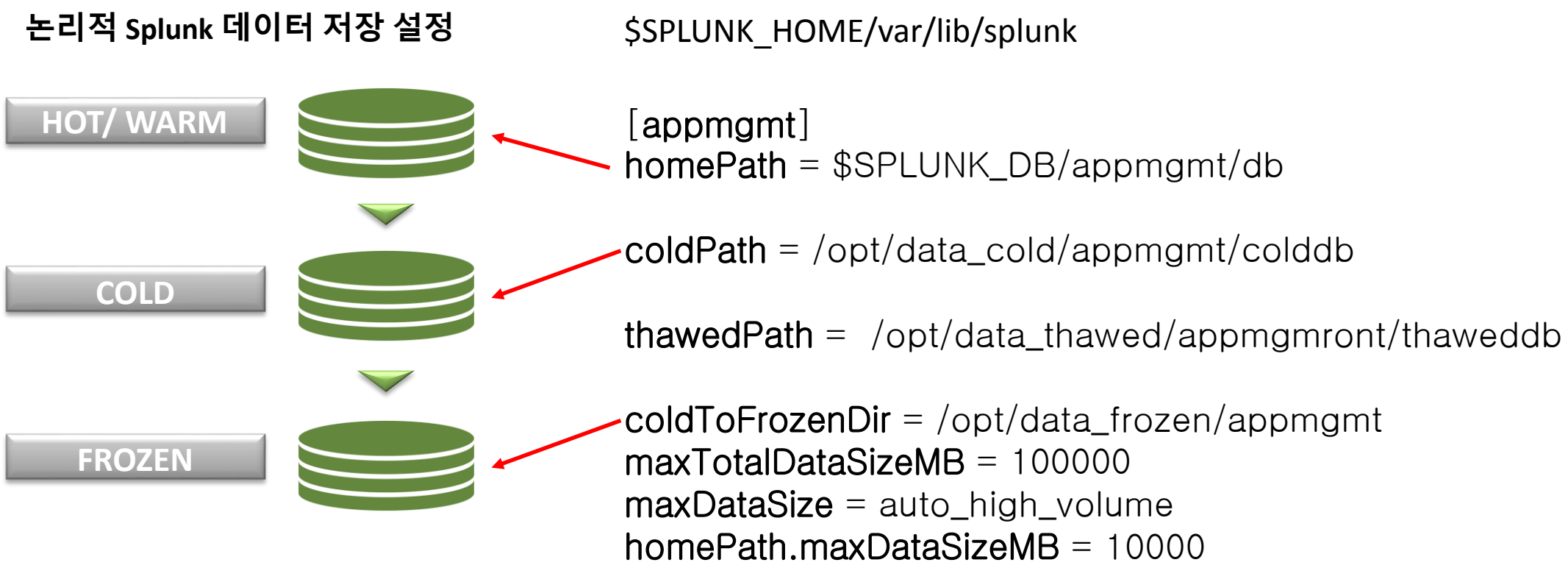
Splunk Indexer 데이터 저장 구조

> **Storage** 에 따른 데이터 관리 전략



Splunk Indexer 데이터 저장 구조

> Indexes.conf 파일 설정



Splunk Forwarder 설명

Forwarder

- **역할,** 데이터 수집 및 전달
- **전송 유형**
  - 원시 데이터
  - 필터링 되지 않은 데이터
  - 필터링 된 데이터

Heavy Forwarder

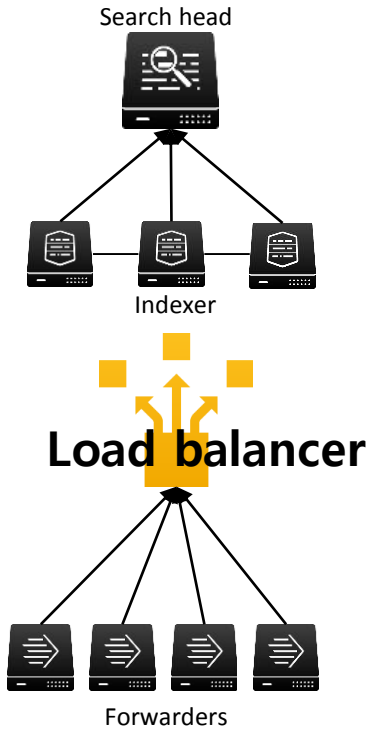
- 원시 데이터 전송
- 필터링 되지 않은 데이터
- 필터링 된 데이터

Universal Forwarder

데이터 전달하는 데 필요한 구성 요소만 포함됨

- CPU에 대한 부하가 적음
- 메모리를 사용이 적음
- 디스크 공간 사용이 작음
- 기본 데이터 전송률은 256Kbps
- 데이터 필터링 및 수정 불가
- Heavy Forwarder 같이 **Python** 번들 버전 없음
- Heavy Forwarder 인스턴스로 변환할 수 없음

Splunk Forwarder 종류

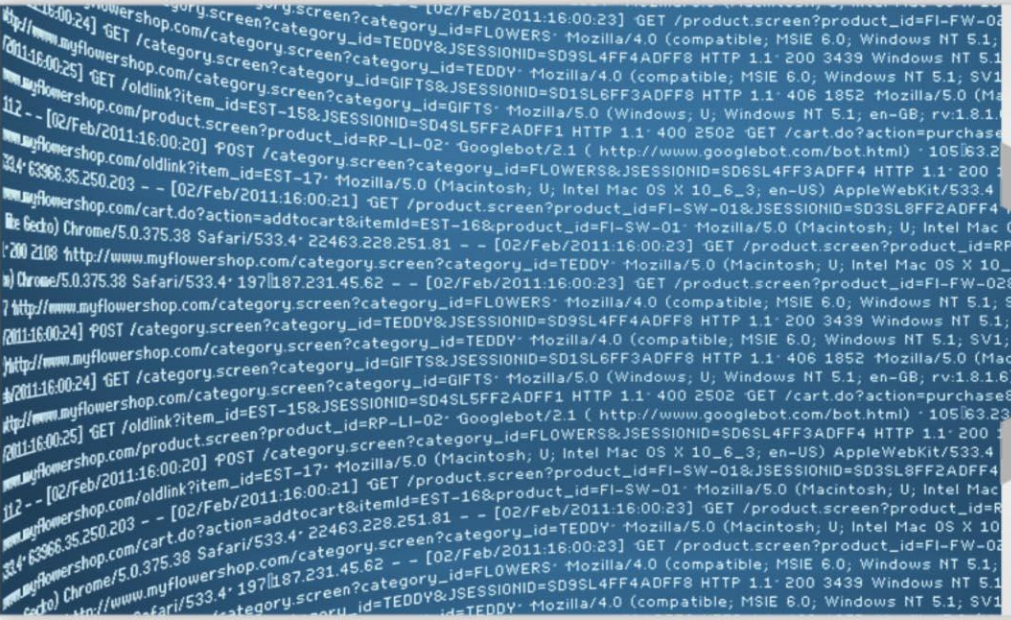


> Load balancing Forwarder

- Indexer로 분산 저장 시 사용
- Indexer와 Forwarder의 수평확장을 용이하게 함
- 데이터의 분산배포는 Forwarder의 인터벌 타임의 조작을 통해 주기의 조절이 가능함

자료 : http://docs.splunk.com/Documentation/Splunk/6.2.1/Forwarding/Setuploadbalancingd

splunk> MOS



Splunk Fields

splunk>

Splunk Metadata

> **Metadata** (Indexing 시 생성되는 필드)

Splunk는 인덱싱 과정에서 각 이벤트에 대해 host, source 및 sourcetype을 포함한 기본 필드를 추출함

필드유형	필드 리스트	설명
내부필드	_raw, _time, _index, _key	▪ Splunk 이벤트에 대한 일반적인 정보를 포함하는 필드
주요 metadata (모든 이벤트에 공통으로 해당됨)	host, index, linecount, punct, source, sourcetype, splunk_server, timestamp	▪ 이 필드는 인덱싱 되고 필드 메뉴에 **기본적**으로 추가됨 ▪ 이벤트가 시작된 장소 ▪ 이벤트가 위치해 있는 인덱스 ▪ 이벤트의 유형 ▪ 이벤트에 포함된 줄 수 ▪ 이벤트가 발생한 시간에 대한 정보가 포함된 필드
시간필드	date_hour, date_mday, date_minute, date_month, date_second, date_wday, date_year, date_zone	▪ 타임스탬프에 더 세부적으로 추가된 정보 필드 ▪ date_* 필드는 각 시스템에서 생성된 타임스탬프 정보가 있는 이벤트만 생성 됨

Metadata

index
- 데이터가 인덱싱 된 인덱스(DB)의 이름이 표시된 필드
- 모든 이벤트는 기본적으로 main 인덱스(index="main")에서 인덱싱 됨
- 검색 예) "index=mosbootcamp*" (와일드카드를 사용 가능)

host
- 네트워크 장치의 호스트 이름 혹은 IP 주소가 표시된 필드
- 검색 예) "host=mos*" (와일드카드를 사용 가능)

source
- 이벤트가 시작된 파일, 스트림 또는 기타 입력의 이름이 표시된 필드
- 데이터 생성 명령의 결과를 필터링 하기 위해 사용하거나 데이터 처리 명령의 인수로 사용 가능
- 검색 예) "source=mossplunk*" (와일드카드를 사용 가능)

Metadata

sourcetype
- 이벤트가 시작된 데이터 입력 형식이 지정된 필드
- 검색 예) "sourcetype=access*" (와일드카드를 사용 가능)

_time
- Unix Time의 이벤트의 타임스탬프를 표시한 필드
- Splunk Web에 이벤트 시간 표시줄 생성

linecount
- 이벤트가 인덱싱 되기 전에 이벤트의 라인의 수가 표시된 필드
- 특정 수의 줄과 일치하는 이벤트를 검색하기 위해 사용하거나 데이터 처리 명령에서 인수로 사용