### 수집 모델

#### 개요

수집 모델은 데이터의 수집부터 정규화에 이르는 일련의 데이터 처리 과정을 정의하는 로드맵과 같습니다. 수집 모델에 원본 데이터를 수집할 수집기가 어떻게 동작해야 하는지, 수집된 데이터를 어떻게 파싱할 것인지, 파싱한 데이터를 어떻게 정규화할 때 필요한 규칙이 모두 정의됩니다.



수집 모델에 따라 파서가 지정되어 있지 않은 경우도 있습니다. 로그 데이터가 이미 구조화된 형태로 되어 있을 경우에는 별도의 파싱 과정이 필요하지 않습니다. 예를 들어, 데이터베이스에 있는 데이터는 이미 필드별로 구분되어 있으므로 별도로 파싱할 필요가 없습니다.

앱 수집 모델과 기본 수집 모델

수집 모델은 기본 수집 모델과 앱 수집 모델로 구분할 수 있습니다. 파서와 마찬가지로 관리자가 수집 모델을 구성해 사용해야 하는 경우는 많지 않습니다.

**기본 수집 모델**

앱을 설치하지 않은 상태에서 로그프레소 소나에서 설정할 수 있는 수집 모델. 초기에는 다음과 같이 두 가지 기본 수집 모델만 제공합니다.

|  |  |
| --- | --- |
| 이름 | 설명 |
| WTMP | wtmp 로그 파일의 수집, 정규화 방법을 정의한 수집 모델 |
| 윈도우 XML 이벤트 로그 | 윈도우의 이벤트 로그를 수집/정규화하는 방법을 정의한 수집 모델 |

윈도우 XML 이벤트 로그 수집 모델은 윈도우 센트리가 있는 환경에서 사용할 수 있습니다.

다음은 로그프레소 소나에서 설정할 수 있는 기본 수집 모델의 유형입니다. 유형에 따라 수집기의 [수집 설정](https://docs.logpresso.com/ko/sonar/4.0/ui/section-loggers) 항목이 결정됩니다.

|  |  |
| --- | --- |
| 수집기 유형 | 설명 |
| CEP 이벤트 | 이벤트 연관 분석(CEP) 기능을 이용하는 데이터 수집 |
| DNS 스니퍼 | PCAP 장치를 이용해 DNS 패킷을 수집 |
| FTP 디렉터리 와처 | FTP 통신으로 원격 호스트의 특정한 디렉터리에서 텍스트 로그 파일을 수집 |
| FTP 로테이션 로그 파일 | FTP 통신으로 원격 호스트에서 로테이션되는 단일 텍스트 로그 파일을 수집 |
| FTP 일자별 디렉터리 | FTP 통신으로 하루 단위로 생성되는 디렉터리를 순회하며 패턴과 이름이 일치하는 텍스트 로그 파일을 수집 |
| GZIP 디렉터리 와처 | 로컬 호스트에서 GZIP 형식으로 압축된 텍스트 로그 파일을 수집 |
| HTTP POST | 원격 호스트가 HTTP POST 메소드로 전송하는 데이터 수집 |
| HTTP 모니터 | 원격 호스트의 웹 서비스 응답 코드와 통계 정보를 수집 |
| HTTP 스니퍼 | PCAP 장치를 이용해 HTTP 패킷을 수집 |
| JDBC 수집기 | JDBC를 이용해 데이터베이스에서 데이터 수집 |
| NetFlow | NetFlow v5/v9 패킷을 수집 |
| PCAP 디렉터리 와처 | 로컬 디렉터리에서 PCAP 파일을 수집 |
| PCAP 패킷 | PCAP 장치를 이용해 모든 패킷을 수집 |
| RSS 수집기 | URL로 지정된 RSS 피드에서 데이터 수집 |
| SFTP WTMP 파일 | SFTP 통신으로 원격 호스트의 WTMP 로그를 수집 |
| SFTP 디렉터리 와처 | SFTP 통신으로 원격 호스트의 특정한 디렉터리에서 텍스트 로그 파일을 수집 |
| SFTP 로테이션 로그 파일 | SFTP 통신으로 원격 호스트에서 로테이션되는 단일 텍스트 로그 파일을 수집 |
| SFTP 일자별 디렉터리 | SFTP 통신으로 하루 단위로 생성되는 디렉터리를 순회하며 패턴과 이름이 일치하는 텍스트 로그 파일을 수집 |
| SNMP GET | SNMP 에이전트에게서 수집한 데이터 수집 |
| SNMP 인터페이스 통계 | SNMP 에이전트의 네트워크 인터페이스별 트래픽 통계 수집 |
| SNMP 트랩 수신기 | SNMP 트랩으로 수신한 SNMP 메시지 수집 |
| SNMPv3 GET | SNMPv3 에이전트에게서 수집한 데이터 수집 |
| SNMPv3 인터페이스 통계 | SNMPv3 에이전트의 네트워크 인터페이스별 트래픽 통계 수집 |
| SSH 실행 | SSH 셸에서 실행한 명령의 표준 출력을 수집 |
| TCP 포트 상태 수집기 | 로컬/원격 호스트의 TCP 포트 개방 여부를 확인하고 결과를 수집 |
| WTMP | 로컬 호스트의 WTMP 로그를 수집 |
| sFlow | sFlow v5 패킷을 수집 |
| 디렉터리 와처 | 로컬 호스트의 특정한 디렉터리에서 텍스트 로그 파일을 일정 주기마다 수집 |
| 디스크 사용량 | 로그프레소 소나 노드 또는 센트리의 디스크 사용량을 주기적으로 모니터링하여 수집 |
| 로테이션 로그 파일 | 로컬 호스트에서 로테이션되는 단일 텍스트 로그 파일을 수집 |
| 리눅스 계정 이벤트 탐지 | 로컬 호스트에서 계정 목록 파일 passwd의 변경사항을 수집 |
| 리커시브 GZIP 디렉터리 와처 | 로컬 호스트의 특정한 디렉터리에서 패턴과 이름이 일치하는 GZIP 로그 파일을 수집 |
| 리커시브 디렉터리 와처 | 로컬 호스트의 지정된 디렉터리에서 패턴과 이름이 일치하는 텍스트 로그 파일을 수집 |
| 멀티 로테이션 로그 파일 | 일정 주기마다 다른 경로에 백업 후 삭제하고 다시 쓰는 로그 파일들을 수집 |
| 설정 파일 변경 탐지 | 로컬 호스트에서 패턴과 이름이 일치하는 파일의 변경사항을 수집 |
| 센트리 성능 로그 | 로그프레소 소나에서 센트리 성능 로그를 일괄 수집 |
| 스트림 쿼리 출력 | 로그프레소 소나에서 스트림 쿼리의 출력 데이터를 수집 |
| 시스로그 수집기 | 원격 호스트가 전송하는 Syslog 메시지를 수집 |
| 외부 프로그램 | 로컬 호스트에서 실행한 명령, 스크립트의 표준 출력을 수집 |
| 웹 키퍼 | MSSQL을 사용하는 구 버전 웹 키퍼의 데이터를 수집 |
| 일자별 디렉터리 | 로컬 호스트에서 하루 단위로 생성되는 디렉터리를 순회하면서 패턴과 이름이 일치하는 텍스트 로그 파일을 수집 |
| 포워더 성능 로그 | 로그프레소 소나에서 전달 노드의 성능 로그를 일괄 수집 |

**앱 수집 모델**

대부분의 수집 모델은 [앱](https://docs.logpresso.com/ko/sonar/4.0/ui/chapter-apps)을 설치할 때 함께 제공됩니다. 앱 수집 모델은 연동 대상 시스템으로부터 수집된 데이터에서 필드를 추출하는데 필요한 모든 설정이 되어 있습니다.

앱 수집 모델을 수정하거나 삭제하지 마세요. 앱의 동작에 문제를 일으킬 수 있습니다.

정규화 규칙

수집 모델은 로그 스키마에 따라 로그를 정규화하는 규칙을 제공합니다. 정규화 규칙은 로그 스키마, 입력된 로그 데이터를 가공하는데 필요한 쿼리문, 새로고침 주기로 구성됩니다.

하나의 로그 데이터에 여러 가지 유형의 로그 레코드가 섞여 있을 수 있으므로 정규화 규칙들은 상호 배타적(mutially exclusive)이고 모든 것을 포괄(collectively exhaustive)할 수 있게 작성되어야 합니다(MECE 원칙). 어떤 로그 데이터에 A, B, C 세 가지 유형의 레코드가 섞여 있다고 하면, 수집 모델은 각각의 유형에 대응하는 세 가지 정규화 규칙을 포함하고 있어야 합니다. 팔로알토 네트웍스 방화벽 로그는 세션 로그와 침입탐지 로그로 구성됩니다. 이렇게 다른 유형의 로그 레코드는 각 레코드에 적합한 로그 스키마에 맞춰 정규화가 필요합니다.

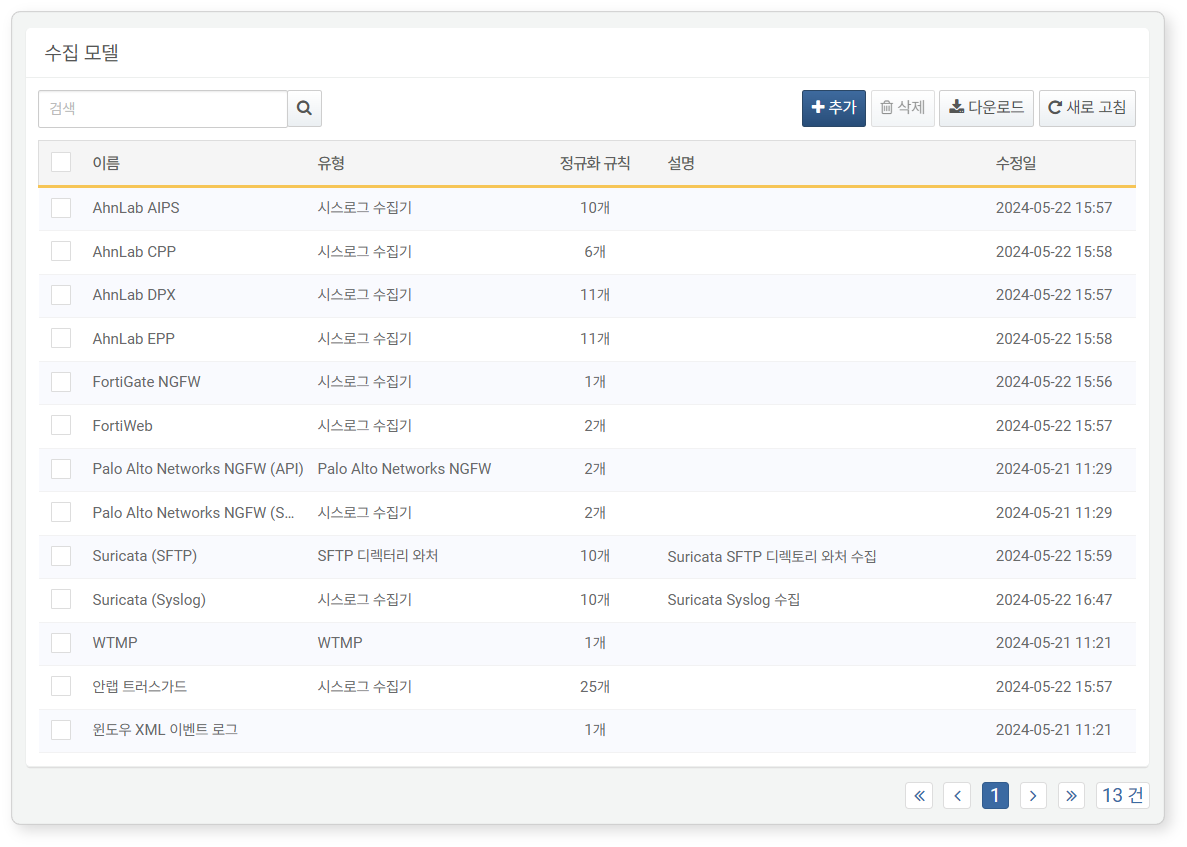
또한 로그 형식의 변경 가능성과 파싱 오류를 감안해 어느 정규화 규칙에도 해당하지 않거나, 정규화할 수 없는 로그 레코드를 보존하는 정규화 규칙도 필요합니다.



제시된 그림은 [Suricata](https://logpresso.store/ko/apps/suricata) 앱의 수집 모델 화면입니다. **Alert**, **Anomaly**, **DHCP**, **DNS**, **Fileinfo**, **Flow**, **HTTP**, **SSH**, **TLS** 등 정규화 규칙 외에 **미분류** 규칙이 있는 것을 확인할 수 있습니다. **미분류** 규칙의 **스트림 쿼리**는 입력으로 받은 데이터 중에서 event\_type 필드의 값이 없거나, 각 정규화 규칙에 정의된 값(alert, anomaly, dhcp, dns, fileinfo, flow, http, ssh, tls)과 일치하지 않는 레코드를 찾아서 미분류로 분류하는 쿼리문입니다.

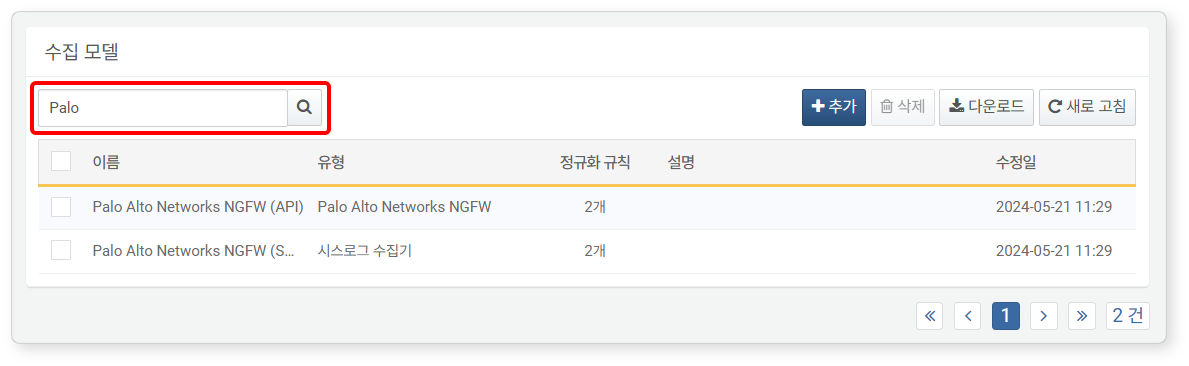
#### 수집 모델 목록 조회/검색

**수집 > 수집 모델**에서 수집 모델 목록을 조회하거나 검색할 수 있습니다.



* **이름**: 수집 모델을 식별에 사용하는 이름
* **유형**: 실행할 수집기의 유형
* **정규화 규칙**: 수집 모델에 정의된 정규화 규칙 개수
* **설명**: 수집 모델에 대한 부가적인 정보
* **수정일**: 수집 모델이 생성된 날짜, 또는 마지막으로 수정한 날짜

수집 모델 목록에서 특정한 수집 모델을 찾으려면 도구 모음에 있는 검색 도구를 사용하세요. 검색 도구는 **이름**에서 입력한 단어가 포함된 수집 모델을 찾아서 보여줍니다. 수집 모델 검색 도구는 대소문자를 구분하지 않습니다.

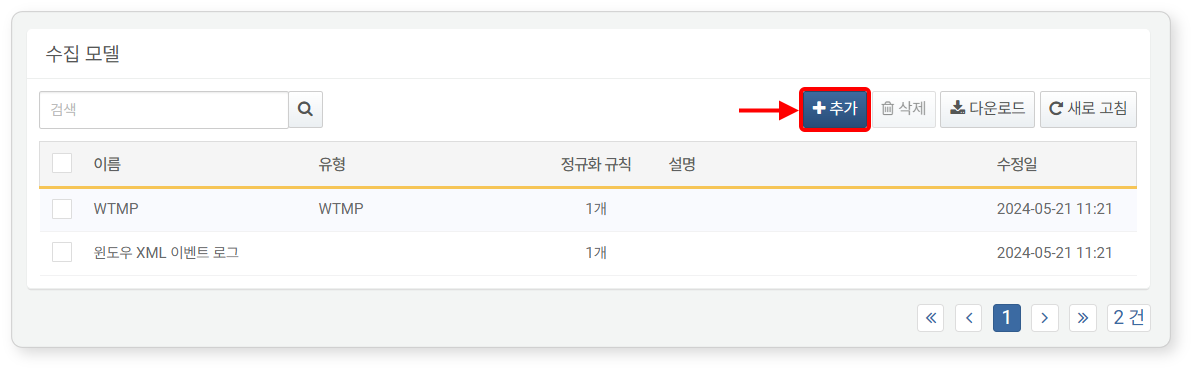


#### 수집 모델 추가

대부분의 운영 환경에서 앱 수집 모델을 이용하는 것으로 충분합니다. 그러나, 새로운 수집 모델이 필요한 경우, 수집 모델을 추가해 사용하세요.

새 수집 모델을 추가하려면,

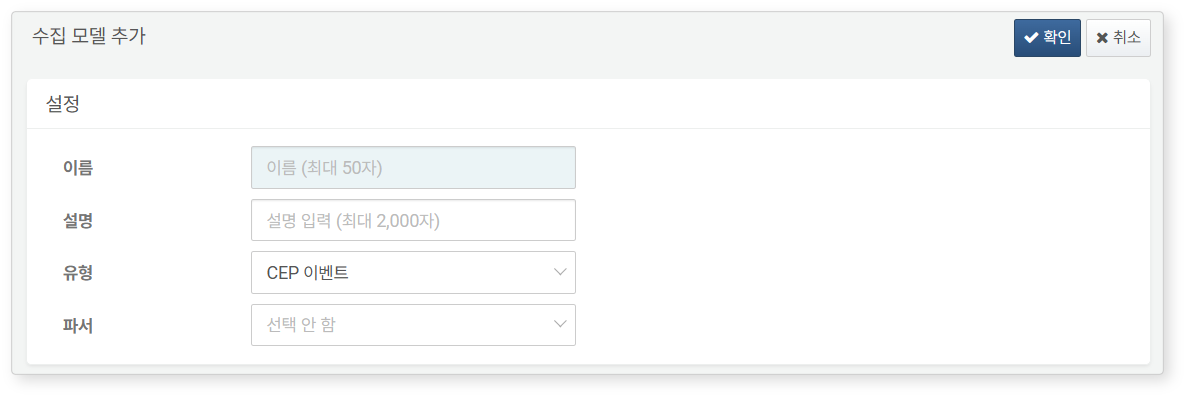
**수집 > 수집 모델**에서 도구 모음에 있는 **추가**를 클릭하세요.



[설정](https://docs.logpresso.com/ko/sonar/4.0/ui/section-logger-models)과 [정규화 규칙](https://docs.logpresso.com/ko/sonar/4.0/ui/section-logger-models)에 필요한 값을 입력하거나 선택한 다음, 우측 상단에 있는 확인을 클릭하세요.

설정

**설정**에서 수집 모델의 **이름**, **설명**, **유형**, **파서**를 입력하거나 선택하세요.

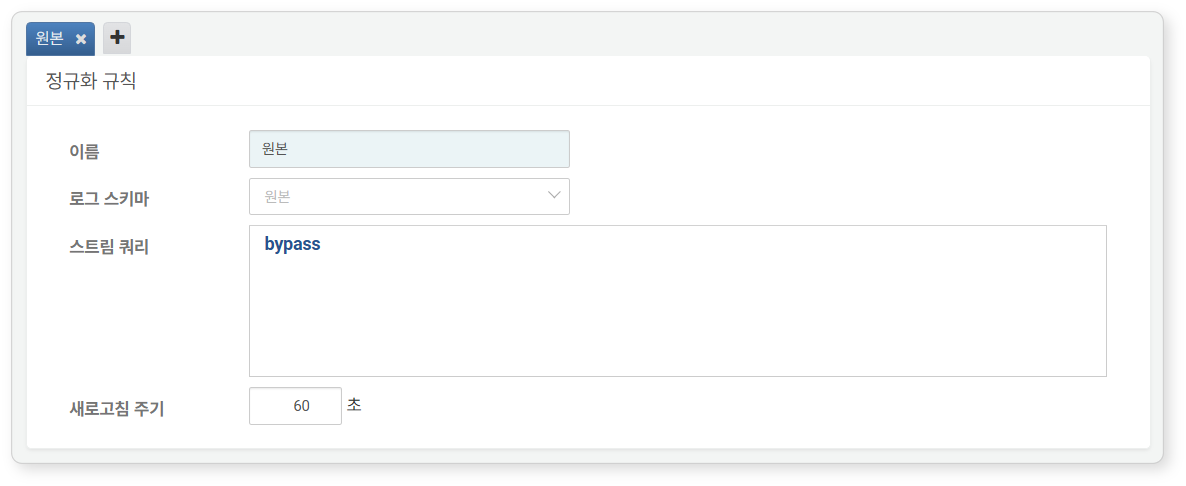


* **이름**: 웹 콘솔에서 사용자가 수집 모델을 식별할 수 있도록 부여하는 이름
* **설명**: 수집 모델에 대한 설명. 수집 모델에 대한 부가적인 정보를 입력할 때 사용하세요.
* **유형**: 실행할 수집기의 유형(기본값: 선택 안 함).
* **파서**: 수집기로부터 전달받은 데이터를 처리할 파서의 이름. 입력 데이터가 이미 구조화되어 있는 경우, 파서를 선택하지 않을 수 있습니다(예: 유형이 JDBC 수집기인 경우).

정규화 규칙

정규화 규칙은 1개 이상 설정할 수 있습니다. 정규화 규칙은 입력 레코드를 서로 중복되지 않게 분류할 수 있어야 하고, 누락되는 레코드가 없도록(mutually exclusive, collectively exhaustive) 작성되어야 합니다.

* 정규화 규칙을 추가하려면 **+** 탭을 클릭하세요.
* 정규화 규칙을 삭제하려면 탭에서 **X**를 클릭하세요.



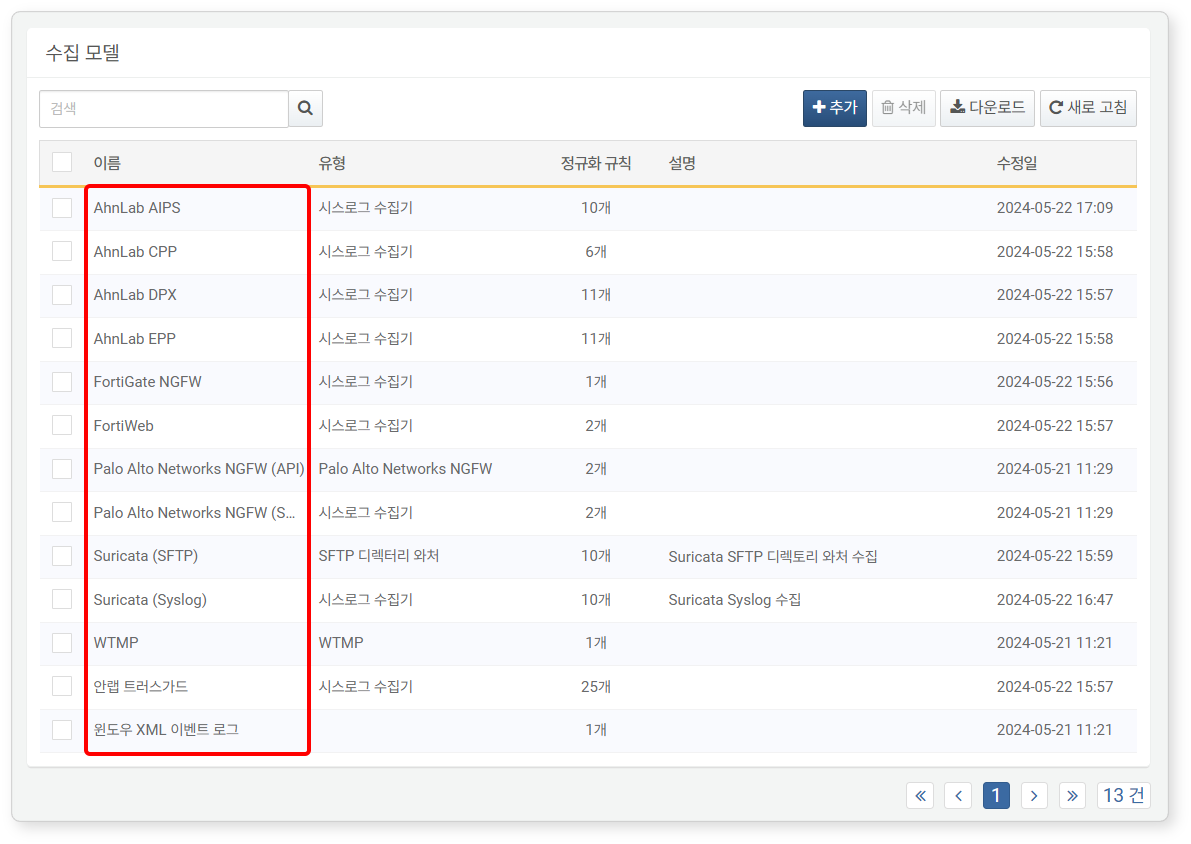
* **이름**: 정규화 규칙을 식별할 수 있는 이름
* **로그 스키마**: 정규화 규칙에 적용할 로그 스키마
* **스트림 쿼리**: 파서가 적용된 데이터를 가공하는데 사용할 로그프레소 쿼리문
* 보통 [search](https://docs.logpresso.com/ko/sonar/4.0/ui/ko/query/search-command) 명령을 이용해 특정한 필드 값을 기준으로 입력 데이터를 필터링는 쿼리문이 사용됩니다.
* 입력 레코드의 필드 이름을 로그 스키마에 맞게 변경하거나, 비어 있는 필드에 값을 할당하는 쿼리문 등을 입력하세요.
* **새로고침 주기**: 입력 데이터의 처리 주기(기본값: 60초, 0으로 지정하면 실시간 처리).

#### 수집 모델 수정

기본/앱 수집 모델에 내장된 정규화 규칙을 수정하지 마세요. 앱의 동작에 문제를 일으킬 수 있습니다.

수집 모델을 수정하려면,

목록에서 수정할 수집 모델의 **이름**을 클릭하세요.



**수집 모델 수정** 화면에서 정보를 수정한 후 **확인**을 클릭하세요.

* 수정할 속성에 대한 설명은 [수집 모델 추가](https://docs.logpresso.com/ko/sonar/4.0/ui/section-logger-models)를 참조하세요.
* **유형**은 수정할 수 없습니다. 유형을 수정하려면 수집 모델을 삭제하고 다시 생성하세요.

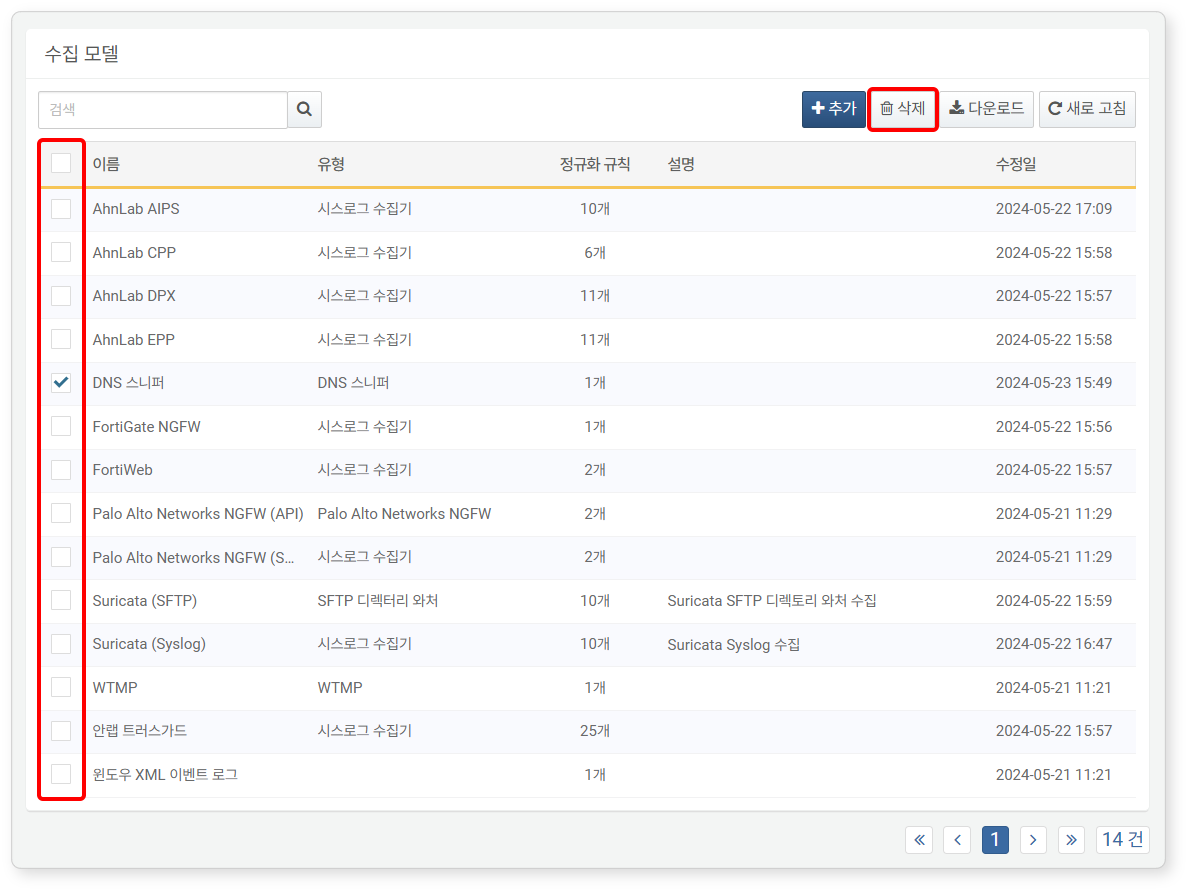
#### 수집 모델 삭제

기본/앱 수집 모델을 삭제하지 마세요. 앱의 동작에 문제를 일으킬 수 있습니다.

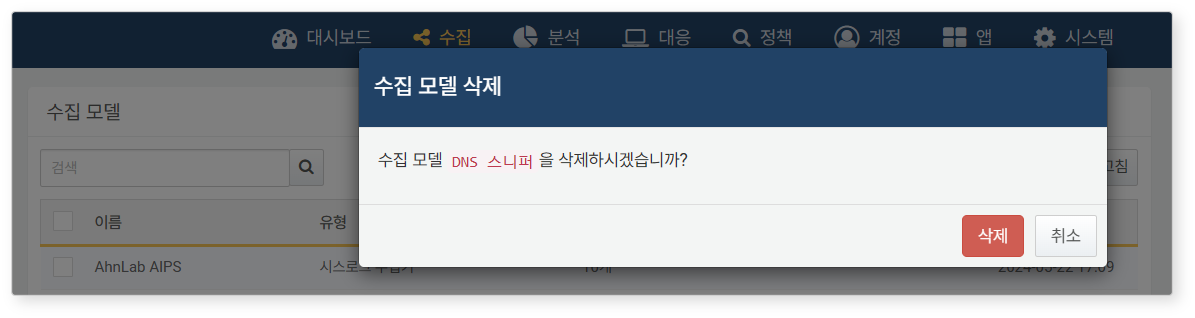
수집 모델을 삭제하려면,

목록에서 삭제할 수집 모델 정보가 있는 행의 체크박스를 선택하세요.

도구 모음에서 **삭제**를 클릭하세요.



**수집 모델 삭제** 대화상자에서 삭제할 수집 모델 목록을 확인한 후 **삭제**를 클릭하세요. 삭제하지 않으려면 **취소**를 클릭하세요.



수집 모델을 삭제하지 못하는 경우, **수집 모델 식제 실패** 대화상자에서 실패 원인을 확인할 수 있습니다.

