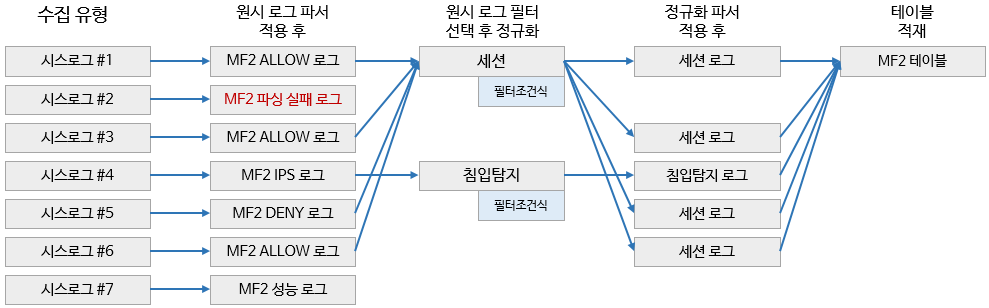
## 수집

로그를 수집하려면 먼저 해당 로그와 관련된 수집 모델이 내장되어 있는지 확인해야 합니다. 내장된 수집 모델로 처리할 수 있는 경우에는 바로 수집 설정으로 가서 수집기를 추가하면 됩니다. 그렇지 않은 경우에는 로그프레소 소나의 데이터 처리 방식을 이해할 필요가 있습니다.



원본 로그가 수집기를 통해 수집 서버로 들어오면 아래의 단계를 거치게 됩니다:

수집기의 수집 모델에 정의된 원본 로그 파서를 이용하여 비정형 로그에서 필드를 추출합니다. DB 테이블에서 레코드를 수집하는 경우처럼 이미 필드가 구분된 경우에는 원본 로그 파서를 지정하지 않습니다.

수집 모델과 연관된 추출 모델의 필터 표현식 조건에 일치하는 경우, 정규화 파서를 적용합니다. 추출 모델에서 필터 표현식이 정의되지 않으면 전체 로그에 대해 정규화 파서를 적용합니다. 정규화 파서는 원본 로그 파서를 통해 추출된 필드를 정규화 스키마의 필드 이름과 타입에 맞게 변환하는 역할을 수행합니다. 추출 모델에서 정규화 파서가 지정되지 않으면 별도로 변환을 수행하지 않습니다.

정규화가 완료된 로그를 테이블에 저장합니다. 이 때 저장되는 레코드는 정규화된 필드 뿐 아니라 원본 로그를 포함하고 있습니다.

이런 일련의 과정은 정규화를 통해 일관된 정책을 수립하고 분석하기 위해 필요합니다. 이기종 장비에서 발생하는 로그는 각각 다른 포맷으로 구성되지만, 공통된 속성을 가지고 있으므로 필드 표준화를 통해서 정책 설정 및 분석을 단순화 할 수 있습니다. 예를 들어, 이기종 방화벽의 로그 포맷은 각각 다르지만, 동일한 필드 이름으로 정규화 된 경우 전체 방화벽을 대상으로 출발지 IP에 대해 블랙리스트 탐지를 지정할 수 있습니다.추출 모델이 정의되지 않으면, 원본 로그는 버려집니다. 파싱에 실패하는 경우에도 원본 로그를 저장하려면 파싱에 실패한 로그를 수신할 수 있도록 필터 조건식을 설정한 추출 모델을 정의합니다. 일반적으로 타입에 해당하는 필드의 존재 여부로 파싱 성공 여부를 확인할 수 있습니다.

새로운 로그를 수집하는 경우, 먼저 기존의 정규화 스키마로 해당 로그를 매핑할 수 있는지 검토합니다. 필요한 경우 정규화 스키마를 새로 정의하고, 수집 모델과 추출 모델을 순서대로 추가합니다. 수집 모델은 로그 수집 방식 (예를 들면 SYSLOG)과 원본 로그 파서의 쌍으로 정의됩니다. 추출 모델은 연관된 수집 모델, 원본 로그 필터, 정규화 파서, 정규화 스키마의 쌍으로 정의됩니다. 모든 준비가 완료되면 수집 설정에서 새 수집기를 추가할 수 있습니다.