### textfile

텍스트 파일에서 데이터를 읽어옵니다.

#### 문법

textfile [OPTIONS] PATH

필수 매개변수

**PATH**

파일 경로. 절대경로로 입력하거나, 로그프레소 엔진을 기준으로 상대경로로 지정할 수 있습니다. 파일 이름에 와일드카드(\*)를 사용해 패턴 매칭 방식으로 파일을 조회할 수 있습니다. 예를 들어, PATH에 allow-\*.txt를 지정함으로써 allow-ip.txt, allow-user.txt, allow-url.txt 등의 파일을 한꺼번에 조회할 수 있습니다. 파일을 읽어오려면 로그프레소 실행 계정에 접근 권한이 부여되어 있어야 합니다.

선택 매개변수

**brex="REGEX"**

하나의 레코드가 여러 행으로 구성된 경우, 레코드의 시작 행을 찾을 때 사용할 정규표현식. 정규표현식이 일치하는 행이 나오기 전 행 혹은 erex로 지정한 마지막 행까지 하나의 레코드로 병합합니다. 지정하지 않으면 개행문자(CRLF 혹은 LF)를 기준으로 인식합니다.

레코드의 마지막 행을 찾으려면 erex 옵션을 이용합니다.

**cs=CHARSET**

문자열 인코딩 형식 (기본값: utf-8). 다음 문서에 등록된 Preferred MIME Name이나 Aliases를 사용합니다: <https://www.iana.org/assignments/character-sets/character-sets.xhtml>

**df="TIME\_FMT"**

dp로 추출한 날짜 데이터를 파싱할 때 사용할 형식 문자열. yyyy-MM-dd HH:mm:ss.SSS와 같이 입력할 수 있습니다. dp 옵션과 함께 사용합니다. 지정하지 않으면 데이터 로딩 시점의 시각을 **\_time** 필드에 기록합니다.

형식 문자열에 다음과 같은 날짜 지시자를 사용할 수 있습니다.

|  |  |  |
| --- | --- | --- |
| 지시자 | 설명 | 예제 |
| G | 기원 전/후 | AD |
| y | 달력 기준 연도 | 2025 (yyyy); 25 (yy) |
| M | 월 | July (MMMM); Jul (MMM); 07 (MM), 7 (M) |
| w | 연 기준 몇 번째 주 | 27 (해당 연도의 27번째 주) |
| W | 월 기준 몇 번째 주 | 2 (해당 월의 두 번째 주) |
| D | 연 단위 일 | 10 |
| d | 월 단위 일 | 189 |
| E | 요일 | Tuesday (EEEE); Tue (E, EE, EEE) |
| F | 월 기준 몇 번째 요일 | 2 (해당 월의 두 번째 요일) |
| u | 요일 숫자(1=월요일, …, 7=일요일) | 1 |
| a | 오전/오후 | PM |
| H | 시간(0-23) | 0 |
| k | 시간(1-24) | 24 |
| K | 오전/오후 시간 (0-11) | 0 |
| h | 오전/오후 시간 (1-12) | 12 |
| m | 분 | 30 |
| s | 초 | 55 |
| S | 밀리초 | 978 |
| z | 시간대(일반 표기) | Pacific Standard Time; PST |
| Z | 시간대(RFC 822 표기) | -0800 |
| X | 시간대(ISO 8601 표기) | -08;-0800;08:00 |

예를 들어, 읽어온 데이터에 2000-01-01 11:22:33과 같은 문자열이 있고, 이와 유사한 문자열에서 날짜 데이터를 추출할 형식 문자열이 yyyy-MM-dd HH:mm:ss 라면 **\_time** 필드에 2000년 1월 1일 11시 22분 33초가 기록됩니다. df를 지정하지 않으면 데이터를 읽어들인 시점의 시각을 **\_time** 필드에 기록합니다.

**dp="REGEX"**

**\_time** 필드에 기록할 날짜 추출 정규표현식. 예를 들어, 읽어온 데이터에 2000-01-01 11:22:33과 같은 문자열이 있고, 이와 동일한 패턴의 문자열에서 날짜 데이터를 추출하려면 dp에 정규표현식으로 (\d{4}-\d{2}-\d{2} \d{2}:\d{2}:\d{2})를 지정합니다. (정규표현식 메타 문자인 \가 문자열 안에 있으므로 이스케이프 문자로서 \를 추가해 \를 메타 문자로 사용합니다.) dp로 추출한 날짜 데이터는 df로 지정한 지정한 형식으로 파싱합니다. df 옵션과 함께 사용합니다. 이 옵션을 지정하지 않으면 데이터를 읽어들인 시점의 시각을 **\_time** 필드에 기록합니다.

**erex="REGEX"**

하나의 레코드가 여러 행으로 구성된 경우, 레코드의 마지막 행을 찾을 때 사용할 정규표현식. 정규표현식이 일치하는 행이 나오기 전까지, 혹은 파일의 마지막 행부터 파일 끝까지 하나의 레코드로 병합합니다. 지정하지 않으면 개행문자(CRLF 혹은 LF)를 기준으로 인식합니다.

레코드의 시작 행을 찾으려면 brex 옵션을 이용합니다.

**limit=INT**

가져올 최대 로그 개수(기본값: 제한 없음)

**offset=INT**

건너뛸 로그 개수(기본값: 0)

#### 사용 예

/var/log/secure 로그 파일 조회

textfile /var/log/secure

euc-kr로 인코딩 된 iis.txt 파일 조회

textfile cs=euc-kr iis.txt

/var/log에서 파일 이름에 syslog.가 포함된 모든 gz 파일 조회

textfile /var/log/syslog.\*.gz

test.txt 파일 내용 중 첫 5줄은 건너뛴 후 20건 조회

textfile offset=5 limit=20 test.txt

레코드 시작 행은 대괄호()로 감싸여진 문자열을 포함하고 마지막 행은 end 문자열을 포함하는 test.txt 파일 조회

textfile brex="\[.\*\]" erex="end" test.txt

test.txt 파일을 조회하면서 레코드 내에 2000-01-01 11:22:33 형태로 포함된 날짜 데이터를 \_time 필드로 추출

textfile dp="(\\d{4}-\\d{2}-\\d{2} \\d{2}:\\d{2}:\\d{2})" df="yyyy-MM-dd HH:mm:ss" test.txt

#### 호환성

읽을 파일의 확장자가 .gz이면 자동으로 gzip 파일로 인식하여 압축 해제 후 조회할 수 있습니다. 이 기능은 ENT #2241 2019-04-23\_17-20 버전부터 지원합니다.