### csvfile

CSV(Comma-Separated Values) 또는 TSV(Tab-Separated Values) 파일에서 데이터를 조회합니다. CSV 또는 TSV 파일 첫번째 줄에 있는 헤더 정보를 읽어와 필드명으로 사용합니다.

#### 문법

csvfile [OPTIONS] PATH

필수 매개변수

**PATH**

CSV 파일의 경로, 또는 파일 경로를 반환하는 쿼리 매개변수를 참조합니다. 파일 이름에 와일드카드(\*)를 사용해 패턴 매칭 방식으로 파일을 조회할 수 있습니다. 예를 들어, PATH에 allow-\*.csv를 지정함으로써 allow-ip.csv, allow-user.csv, allow-url.csv 등의 파일을 한꺼번에 조회할 수 있습니다. 파일을 읽어오려면 로그프레소 실행 계정에 접근 권한이 부여되어 있어야 합니다.

선택 매개변수

**cs=CHARSET**

문자열 인코딩 형식(기본값: utf-8). 이 옵션은 대소문자를 구분하지 않으며, 다음 문서에 등록된 Preferred MIME Name이나 Aliases를 CHARSET으로 사용할 수 있습니다: <http://www.iana.org/assignments/character-sets/character-sets.xhtml>

**limit=INT**

가져올 레코드의 최대 개수(기본값: 제한 없음)

**maxcol=INT**

조회할 최대 컬럼 수(기본값: 10,000 개). 조회할 데이터의 컬럼 개수가 지정한 최대 컬럼을 초과하는 경우, rest 옵션을 이용해 처리 방식을 정의합니다.

**offset=INT**

건너뛸 레코드 개수(기본값: 0)

**rest=BOOL**

maxcol 옵션으로 지정한 개수 이후의 컬럼의 표시 여부(기본값: f).

1. t: maxcol 옵션으로 지정한 개수를 초과하는 컬럼의 데이터를 모두 **\_rest** 필드에 할당
2. f: 지정한 개수를 초과하는 컬럼의 데이터를 모두 버림

**strict=BOOL**

RFC4180[(https://tools.ietf.org/html/rfc4180)](https://tools.ietf.org/html/rfc4180)의 준수 옵션 (default: f)

1. t: 마이크로소프트 엑셀과 동일하게 RFC4180 준수. 이 옵션을 tab=t와 함께 사용할 수 없습니다.
2. f: CSV 파일을 유연하게 파싱

**tab=BOOL**

탭(tab) 문자를 구분자로 사용 여부 (기본값: f).

1. t: 탭(tab) 문자를 구분자로 사용
2. f: 쉼표(,)를 구분자로 사용

#### 사용 예

/opt/logpresso/wp-nginx.csv 파일을 조회

# 다운로드: https://raw.githubusercontent.com/logpresso/dataset/main/wp-nginx.csv | csvfile /opt/logpresso/wp-nginx.csv

/opt/logpresso/wp-nginx.csv 파일 내용 중 첫 줄을 건너뛴 후 20건의 레코드 조회

csvfile limit=20 offset=1 /opt/logpresso/wp-nginx.csv

/opt/logpresso/wp-nginx.csv 파일에서 4개의 컬럼만 조회

csvfile maxcol=4 /opt/logpresso/wp-nginx.csv

/opt/logpresso/wp-nginx.csv 파일에서 4개의 컬럼만 조회하고, 나머지 컬럼 값을 **\_rest** 필드에 할당

csvfile maxcol=4 rest=t /opt/logpresso/wp-nginx.csv

구분자와 컬럼 사이에 공백이 있는 데이터

strict=t일 때, 구분자와 컬럼 사이에 공백이 있으면 큰 따옴표(")는 문자로 인식되어 의도한대로 파싱되지 않습니다.

# 다운로드: https://raw.githubusercontent.com/logpresso/dataset/main/csvfile-strict-option-test-1.csv | csvfile strict=t /opt/logpresso/csvfile-strict-option-test-1.csv

strict=f일 때는 큰 따옴표 쌍(" ")의 짝이 맞으면 따옴표 쌍 안에 있는 문자열만 컬럼으로 인식하기 때문에 의도한대로 파싱됩니다.

csvfile strict=f /opt/logpresso/csvfile-strict-option-test-1.csv

구분자와 컬럼 사이에 공백이 없는 데이터

strict 값에 관계없이 구분자와 컬럼 사이에 공백이 없으므로 의도한대로 파싱됩니다.

# 다운로드: https://raw.githubusercontent.com/logpresso/dataset/main/csvfile-strict-option-test-2.csv | csvfile strict=t /opt/logpresso/csvfile-strict-option-test-2.csv csvfile strict=f /opt/logpresso/csvfile-strict-option-test-2.csv

큰 따옴표(")를 이스케이프 처리(")한 데이터

strict=t일 때, 이스케이프 문자(\)를 일반 문자로 인식하므로 큰 따옴표 쌍(" ")으로 감싼 컬럼 안에서 "를 사용하여 큰 따옴표(")를 표기한 경우, 의도한대로 파싱되지 않습니다.

# 다운로드: https://raw.githubusercontent.com/logpresso/dataset/main/csvfile-strict-option-test-3.csv | csvfile strict=t /opt/logpresso/csvfile-strict-option-test-3.csv

strict=f일 때, 겹 큰 따옴표("")와 이스케이프 처리한 큰 따옴표(")가 의도한대로 컬럼 안에서 큰 따옴표로 파싱됩니다.

csvfile strict=f /opt/logpresso/csvfile-strict-option-test-1.csv csvfile strict=f /opt/logpresso/csvfile-strict-option-test-3.csv

#### 호환성

maxcol과 rest 옵션은 ENT #2246 2019-05-24\_14-58 버전부터 지원합니다.