

D1 읽기 한도 때문에 방문 통계를 「한 번 읽고 코드에서 묶기」로 바꾼 일

작성 2026-10-20 · 읽는 시간 약 5분 · 작성 루리의 줍줍 편집

#Cloudflare #방문 집계

Cloudflare D1(서버리스 SQLite)을 무료 한도로 쓰는 사이트예요. 관리자 화면의 방문 통계를 열 때마다 읽기 한도가 눈에 띄게 줄어서, 통계 한 장을 보는 데 한도를 거의 다 쓰는 일이 생겼어요. 원인은 쿼리 하나하나가 비싼 게 아니라 **같은 테이블을 열세 번 훑는 구조**였어요. 이 글은 그걸 「한 번 읽고 코드에서 묶기」로 바꾼 과정이에요. 한도의 정확한 숫자는 플랜과 시점에 따라 달라서 적지 않았고, 아래 구조는 이 사이트 워커 코드에서 가져온 거예요.

문제: 통계 화면 한 번에 읽기 한도가 바닥

D1의 읽기 한도는 「읽은 행 수」 기준이에요. 쿼리 수가 아니라 **쿼리가 훑은 행 수**가 쌓이죠. 방문 기록 테이블(hits)에 행이 수천 개 쌓이면서, 통계 화면이 오늘 방문, 7일 합계, 유입 채널, 인기 페이지, 국가별 같은 값을 각각 COUNT(DISTINCT ...).GROUP BY 쿼리로 따로 구했어요. 값마다 같은 행들을 처음부터 다시 훑으니, 열세 번이면 한 번 볼 때 읽는 행이 열세 배가 돼요.

원인

- 같은 기간(최근 7일)의 행을 값마다 SQL로 다시 집계했어요.
- 누적 합계를 매번 전체 기록에서 다시 켜어요.
- 결과를 어디에도 저장하지 않고 요청마다 새로 계산했어요.

해결: 한 번 읽고, 요약표와 캐시를 쓰기

1. **최근 7일 행만 한 번 읽는다.** 필요한 컬럼만 이름을 적어 읽고, 나머지 집계(방문자 수, 채널, 인기 페이지, 국가)는 워커 메모리에서 Map·Set으로 묶어요.
2. **누적은 일별 요약표에서 읽는다.** 하루가 지난 날은 요약 행(날짜·방문자 수·조회수) 하나로 저장해 두고, 최근 14일만 다시 계산해서 덮어써요. 전체 기록은 더 이상 훑지 않아요.
3. **결과를 3분 캐시한다.** 계산한 결과 JSON을 키-값 테이블에 만료 시각과 함께 저장하고, 새로고침해도 캐시가 살아 있으면 DB 읽기 없이 돌려줘요. 정말 최신 값이 필요할 때만 ?fresh=1을 붙여요.

```
// 최근 7일 행을 한 번만 읽는다(컬럼 이름을 명시 - 나중에 컬럼이 늘어도 영향 없음)
const rows = (await db.prepare(
  'SELECT day, ts, vid, uid, path, ref, country, city, mobile FROM hits WHERE day >= ?'
).bind(weekAgo).all()).results;

// 방문자 수·채널은 메모리에서 묶는다
const uv = a => new Set(a.map(x => x.vid)).size;
const byDay = new Map();
for (const h of rows) {
  const o = byDay.get(h.day) || { vs: new Set(), pv: 0 };
  o.vs.add(h.vid); o.pv++; byDay.set(h.day, o);
}

// 누적: 저장된 일별 요약 + 이번에 다시 센 최근 날짜로 덮어쓰기
const total = new Map(rollupRows.map(r => [r.day, { v: r.v, pv: r.pv }]]);
for (const [day, o] of byDay) total.set(day, { v: o.vs.size, pv: o.pv });

// 결과는 만료 시각과 함께 저장(다음 3분은 DB 읽기 0)
await db.prepare('INSERT INTO kvcache (k, v, exp) VALUES (?, ?, ?) ON CONFLICT(k) DO UPDATE SET v =
  .bind('visits', JSON.stringify(out), nowSec + 180).run();
```

이 구조가 나중에 도움이 된 이유

얼마 뒤 방문 기록 테이블에 컬럼을 여섯 개 더했는데(다른 글 「방문 집계에 우리 점검 방문이 섞였던 일」 참고), 이 통계 코드는 읽을 컬럼 이름을 직접 적고 있어서 한 줄도 바꾸지 않아도 됐어요. SELECT *를 썼다면 읽는 데이터 양이 컬럼 수만큼 늘었을 거예요.

교훈

- 무료 한도형 DB에서는 **쿼리 개수보다 높은 행 수**를 줄이세요. 같은 행을 여러 번 높은 구조가 가장 비싸요.
- 집계가 정말 SQL이어야 하는지 따져 보세요. 한 번 읽은 몇천 행을 코드에서 묶는 편이 훨씬 싸요.
- 지난 기간은 요약 행 하나로 저장하고, 최근 기간만 다시 계산하세요.
- 사람이 자주 여는 화면에는 짧은 캐시를 두세요. 3분만 캐시해도 새로고침 폭주가 사라져요.
- 쿼리는 컬럼 이름을 명시하세요. 나중에 컬럼을 추가해도 안전하고 읽는 양도 안 늘어요.

English summary

An admin statistics view on Cloudflare D1 (free tier) ate the daily read budget because it scanned the same hits table thirteen times; D1 bills by rows read, not by queries. The fix reads the last seven days once with explicit column names, groups visitors, channels and pages in

Worker memory, keeps older days as one summary row per day, and caches the JSON result for three minutes with a ?fresh=1 escape hatch. Explicit column names later let new columns be added without touching this code.

관련 글

방문 집계에 우리 점검 방문이 섞여 수치가 부풀었던 일 — nohit 플래그와 데이터센터 표시로 고친 과정

하루 방문 214명 중 사람일 가능성이 큰 쪽은 76명이었어요. 링크 확인 방문과 데이터센터 점검 트래픽을 가려낸 규칙, ? nohit=1 플래그, Cloudflare Workers의 ASN·접속 회사 표시, 읽음 신호까지 정리했어요.

← 기술 블로그 홈 · 사이트 제작 로그 · RSS

관련: 개발 지식 전체

이 글에 대해 묻고 싶으면 ? 개발 FAQ에 남겨 주세요.



공유

공유하기

카카오톡



링크 복사

네이버 블로그

밴드

X

Facebook

출처: 루리의 줍줍(sungprofile.cloud) · 최신 내용은 사이트에서 확인하세요.