Engineer Дані

30 днів даних про електроенергію за 180 мс

Розробила варіанти погодинної агрегації споживання електроенергії на Python, SQL і Bash + Jq; під час вимірювання Python обробив один 30-денний набір даних JSONL за 180 мс.

2025

180 ms

на агрегацію місяця показань лічильників

Ситуація. Компанія обробляє великі обсяги різнорідних даних про електроенергію з кількох джерел, кожне з яких має власну частоту звітності — від погодинних до 15-хвилинних інтервалів, а подекуди й з нерегулярними часовими мітками. Через таку мінливість важко скласти цілісний, порівнянний набір даних. Щоб енергетична аналітика була точною, ці дані потрібно нормалізувати до узгоджених погодинних значень споживання, агрегованих за зонами.

Завдання. Мета полягала в тому, щоб узяти сирі дані про події з історичного набору даних (у форматі JSONL) і перетворити їх на погодинно вирівняні показники споживання електроенергії, по одному рядку на кожну годину й зону. Зокрема, завдання охоплювало:

- Вирівнювання даних із часовими мітками до чітких погодинних інтервалів,
- Агрегування загального виробництва електроенергії в межах кожного інтервалу підсумовуванням значень міксу,
- Урахування транскордонного обміну: додавання імпорту й віднімання експорту,
- Виведення підсумкових значень у структурованому, масштабованому форматі, придатному для подальшого аналізу.
- Рішення також мало бути достатньо ефективним, щоб масштабуватися на великі часові вікна (30+ днів) і кілька країн/зон.

Дія. Для цього завдання реалізовано три варіанти рішення — на Python, JQ (для обробки JSON у командному рядку) та SQL, — кожен оптимізований під свій контекст:

Python обрали за гнучкість, зручність роботи з даними та здатність ефективно виконувати трансформації в пам'яті. На функціях чистого Python побудовано пайплайн, який:

- Парсив файли JSONL у структуровані датафрейми
- Зводив часові ряди до погодинних інтервалів
- Агрегував виробництво та обчислював чисте споживання електроенергії (виробництво + імпорт – експорт)
- Експортував результати у форматі CSV або завантажував їх у легку базу даних SQLite для перевірки.

На JQ створено швидке рішення з мінімальними залежностями для середовищ командного рядка — ланцюжок фільтрів JQ.

У PostgreSQL дані JSONL було імпортовано, створено нормалізовані таблиці та написано низку SQL-запитів.

Для оцінки продуктивності кожне рішення було протестовано на історичних наборах даних за 1 день і за 30 днів.

Результат. Реалізація на Python виявилася найшвидшим і найкраще масштабованим рішенням і обробила:

- дані за 1 день лише за 34 мс
- набір даних за 30 днів — за 180 мс

Це підтвердило, що вона придатна для довших часових вікон і водночас вкладається в частки секунди. До того ж її код був зрозумілим і зручним у супроводі, тож його легко було розширити на обробку кількох зон або інтегрувати в ETL-пайплайн.

Загалом підхід із кількома інструментами засвідчив гнучкість у виборі засобів, ґрунтовну оптимізацію продуктивності та надійне розв'язання задач вирівнювання за часом і агрегування в реальних пайплайнах даних

про електроенергію.

Застосуйте цей досвід у своєму проєкті

Використайте «30 днів даних про електроенергію за 180 мс» як відправну точку для схожого завдання у вашій організації. Послуги, що стоять за цією роботою, доступні й для вашого проєкту. Пошук запитів і відсутніх індексів у системі, що рік за роком важчає, та виправлення тих, що справді мають значення.

Переглянути наші послуги Зв'язатися з нами →

Дізнайтеся, що привело до результату

Прочитайте про ситуацію, рішення та результат цього прикладу. Ми можемо обговорити, що відповідає вашим умовам, а що потребує іншого підходу. 30 днів даних про електроенергію за 180 мс — 180 ms на агрегацію місяця показань лічильників

Переглянути наші послуги Зв'язатися з нами →

Потрібні швидкі та ефективні пайплайни даних?

Обговоріть свій проєкт з інженером, відповідальним за виконання. Обсяг, основу розрахунку вартості та передачу результатів узгоджуємо письмово до початку робіт.

Роль у цій роботі: Інженер платформи та даних. Тут задіяні Бази даних, PostgreSQL, SQL, Інженерія даних, Пайплайни даних (ETL/ELT), Аналітика даних, Автоматизація та CI/CD, Оптимізація продуктивності та Python. І, до речі, ми добре вміємо Оптимізація продуктивності баз даних, Розробка пайплайнів даних (ETL/ELT) та Аналітика даних та BI-дашборди.

Зв'язатися з нами Ознайомтеся з нашими послугами

Одна частина **довшого переліку** — і він увесь на цьому сайті.

Кожен запис написано однаково: ситуація, завдання, що ми зробили і що змінилося.

Переглянути весь перелік

Визначте потрібне покращення

Надайте репрезентативний набір даних, запитання до нього та поточне вузьке місце. Узгодимо вимірювання якості, пропускної здатності та відновлення.

Дивитися, скільки це коштує Зв'язатися з нами →

Потрібні швидкі та ефективні пайплайни даних?

Запитайте про рішення в роботі «30 днів даних про електроенергію за 180 мс». Ви говоритимете з людьми, які можуть пояснити роботу й оцінити ваш проєкт.

Переглянути наші послуги

welcome@engineer.company +45 50 10 61 84 Копенгаген, Данія

Візьміть цю сторінку із собою

Візитівка

Дбайливо зроблено з любов'ю — D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Авторське право © 2025 – дотепер · Engineer ApS · Усі права захищено.

Надруковано з https://data.engineer.company/uk/portfolio/a-month-of-meter-readings-in-180-milliseconds/