Engineer Data

30 dages eldata på 180 ms

Har udviklet varianter til timeaggregering af elforbrug i Python, SQL og Bash + Jq; i benchmarken behandlede Python ét JSONL-datasæt for 30 dage på 180 ms.

2025

180 ms

til at aggregere en måneds måleraflæsninger

Situation. Virksomheden behandler store mængder heterogene eldata fra flere datakilder, som hver har sin egen rapporteringsfrekvens — fra timeintervaller til intervaller på 15 minutter og i nogle tilfælde uregelmæssige tidsstempler. Den variation gør det svært at skabe et sammenhængende og sammenligneligt datasæt. For at understøtte præcise energianalyser skal dataene normaliseres til ensartede forbrugsværdier pr. time, aggregeret pr. zone.

Opgave. Målet var at tage rå hændelsesdata fra et historisk datasæt (i JSONL-format) og omdanne dem til elforbrugstal afstemt til hele timer, struktureret som én række pr. time og pr. zone. Konkret gik opgaven ud på at:

- afstemme tidsstemplede data til faste timeintervaller,
- aggregere den samlede elproduktion i hvert interval ved at summere værdierne for produktionsmikset,
- indregne den grænseoverskridende udveksling ved at lægge import til og trække eksport fra,
- skrive de endelige værdier ud i et struktureret, skalerbart format, der egner sig til videre analyse.
- Løsningen skulle desuden være effektiv nok til at skalere til store tidsvinduer (30+ dage) og til flere lande/zoner.

Handling. Opgaven blev grebet an med tre forskellige varianter af løsningen, implementeret i henholdsvis Python, JQ (til JSON-behandling på kommandolinjen) og SQL og hver optimeret til sin egen sammenhæng:

Python blev valgt for sin fleksibilitet, den lette databehandling og evnen til at håndtere transformationer i hukommelsen effektivt. Med rene Python-funktioner blev der bygget en pipeline, som:

- parsede JSONL-filerne til strukturerede dataframes
- samplede tidsseriedataene om til timeintervaller
- aggregerede produktionen og beregnede nettoelforbruget (produktion + import – eksport)
- eksporterede resultaterne som CSV eller indlæste dem i en let SQLite-database, så de kunne efterses.

JQ blev brugt til en hurtig løsning med minimale afhængigheder til kommandolinjemiljøer, bygget som en kæde af JQ-filtre.

I PostgreSQL blev JSONL-dataene importeret, normaliserede tabeller oprettet og en række SQL-forespørgsler skrevet.

For at vurdere ydeevnen blev hver løsning benchmarket med historiske datasæt på både 1 dag og 30 dage.

Resultat. Python-implementeringen viste sig at være den hurtigste og mest skalerbare løsning og klarede:

- 1 dags data på blot 34 ms
- 30 dages datasæt på 180 ms

Det bekræftede, at den egnede sig til længere tidsrum med behandlingstider under et sekund. Den gav også klar kode, der var nem at vedligeholde og let kunne udvides til behandling af flere zoner eller integreres i en ETL-pipeline.

Samlet set viste tilgangen med flere værktøjer fleksibilitet i valget af værktøjer, stærk ydeevneoptimering og robust håndtering af udfordringerne med tidsafstemning og aggregering i rigtige pipelines for eldata.

Brug erfaringen i jeres projekt

Brug »30 dages eldata på 180 ms« som udgangspunkt for en lignende udfordring i jeres organisation. Ydelserne bag arbejdet kan også bruges i jeres projekt. At finde de forespørgsler og manglende indeks bag et system, der er blevet tungere hvert år, og rette dem, der betyder noget.

Se vores ydelser Kontakt os →

Se bag resultatet

Læs om situationen, beslutningerne og resultatet i eksemplet. Vi kan drøfte, hvad der passer til jeres vilkår, og hvad der kræver en anden tilgang. 30 dages eldata på 180 ms — 180 ms til at aggregere en måneds måleraflæsninger

Se vores ydelser Kontakt os →

Brug for hurtige, effektive datapipelines?

Tal om jeres projekt med den ingeniør, der har ansvaret for leverancen. Vi aftaler omfang, prisgrundlag og overdragelse skriftligt, før arbejdet begynder.

Rolle i opgaven: Platform- og dataingeniør. Det trækker på Databaser, PostgreSQL, SQL, Data engineering, Datapipelines (ETL/ELT), Dataanalyse, Automatisering & CI/CD, Optimering af ydeevne og Python. Og når I alligevel er her, er vi ret dygtige til Databaseoptimering, Udvikling af datapipelines (ETL/ELT) og Dataanalyse & BI-dashboards.

Kontakt os Se vores ydelser

Ét stykke af **en længere fortegnelse** — og det hele ligger her på sitet.

Hver post er skrevet på samme måde: situationen, opgaven, hvad vi gjorde, og hvad der ændrede sig.

Se hele fortegnelsen

Definér den ønskede forbedring

Medbring et repræsentativt datasæt, de spørgsmål det skal besvare, og den aktuelle flaskehals. Vi kan aftale målinger af kvalitet, gennemløb og gendannelse.

Se, hvad det koster Kontakt os →

Brug for hurtige, effektive datapipelines?

Spørg til beslutningerne bag »30 dages eldata på 180 ms«. I taler med mennesker, der kan forklare arbejdet og vurdere jeres eget projekt.

Se vores ydelser

welcome@engineer.company +45 50 10 61 84 København, Danmark

Tag denne side med dig

Visitkort

Lavet med kærlighed af D.

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

Copyright © 2025 – nu · Engineer ApS · Alle rettigheder forbeholdes.

Udskrevet fra https://data.engineer.company/da/portfolio/a-month-of-meter-readings-in-180-milliseconds/