Engineer Data

Har designet og drevet 6 ETL/ELT-pipelines

Har designet, implementeret og administreret 6 ETL/ELT-pipelines med Google BigQuery, MSSQL, PostgreSQL, shell-scripting, PL/pgSQL og Transact-SQL og integreret data, så et Python-API kan behandle dem effektivt.

2022 – 2024

6

ETL-pipelines på tre databasemotorer

Situation. Organisationen havde brug for at samle og behandle store mængder tabeldata fra et data warehouse på en fjernserver bag en IPsec VPN. Dataene var afgørende for interne analysedashboards og for eksterne Python-API'er, som kunder og partnere brugte.

Opgave. Målet var at designe, implementere og vedligeholde et sæt robuste, automatiserede ETL/ELT-pipelines, der sikkert kunne hente, transformere og indlæse data i Google BigQuery og samtidig sikre datanøjagtighed, ydeevne og skalerbarhed i alle systemerne.

Handling. Der blev designet, implementeret og administreret 6 ETL/ELT-pipelines fra ende til anden, som omfattede en række teknologier, herunder Google BigQuery, MSSQL, PostgreSQL, shell-scripting, PL/pgSQL og Transact-SQL.

- Der blev etableret sikre forbindelser til en fjernserver beskyttet af en IPsec VPN, så datahentningen kunne automatiseres.
- Download af komprimerede dataarkiver med Parquet-, CSV- og .bak-filer blev planlagt og afviklet.
- Der blev udviklet Bash- og Python-scripts, som udpakkede filerne og klassificerede dem efter type og skema.

- For .bak-filernes vedkommende blev sikkerhedskopierne gendannet i en lokal MSSQL Server-instans med RESTORE DATABASE-arbejdsgange, og skemaintegriteten blev valideret.
- Tabeldata blev indlæst i staging-skemaer i MSSQL og PostgreSQL med værktøjerne bcp, psql og SSIS, afhængigt af kildeformatet.
- Der blev skrevet modulære, genanvendelige PL/pgSQL- og T-SQL-procedurer, som rensede, normaliserede og berigede dataene ud fra forretningslogikken.
- Kuraterede datasæt og tabeller blev eksporteret til mellemformater, komprimeret med gzip og overført sikkert til en Google Cloud Storage-bucket.
- Upload og skemamapping til Google BigQuery blev automatiseret med bq CLI og Python-baserede scripts til dataindlæsning.

Resultat. De automatiserede pipelines reducerede det manuelle arbejde og behandlingstiden markant — fra over 3 timers manuelt arbejde til under 20 minutter fra ende til anden — og dataene blev synkroniseret dagligt i stedet for ugentligt. Python-API'erne, der brugte dataene, fik 30 % bedre ydeevne, og det bedre overblik hjalp forretningsanalytikerne med hurtigere at levere indsigt til interessenterne. Løsningen er fortsat skalerbar og kan udvides med nye datakilder, efterhånden som forretningen vokser.

Brug erfaringen i jeres projekt

Brug »Har designet og drevet 6 ETL/ELT-pipelines« som udgangspunkt for en lignende udfordring i jeres organisation. Ydelserne bag arbejdet kan også bruges i jeres projekt. En struktur, der passer til, hvordan forretningen faktisk fungerer, så de svar, I får brug for senere, stadig kan hentes.

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. Har designet og drevet 6 ETL/ELT-pipelines — 6 ETL-pipelines på tre databasemotorer

Se vores ydelser Kontakt os →

Skal data fra mange kilder samles ét sted?

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: Dataingeniør. Det trækker på Databaser, PostgreSQL, SQL, Data warehousing, Data engineering, Datapipelines (ETL/ELT), Cloud, Automatisering & CI/CD, Netværk & VPN og Python. Og når I alligevel er her, er vi ret dygtige til Backend- & API-udvikling, Databasedesign & datamodellering, Udvikling af datapipelines (ETL/ELT), Design af data warehouse og Klargøring af data til AI.

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 →

Skal data fra mange kilder samles ét sted?

Spørg til beslutningerne bag »Har designet og drevet 6 ETL/ELT-pipelines«. 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/designed-and-ran-6-etl-elt-pipelines/