Engineer Data

Designet og driftet 6 ETL/ELT-pipelines

Designet, implementerte og administrerte 6 ETL/ELT-pipelines med Google BigQuery, MSSQL, PostgreSQL, Shell-skripting, PL/pgSQL og Transact-SQL, og integrerte data for effektiv behandling i et Python-API.

2022 – 2024

6

ETL-pipelines på tvers av tre databasemotorer

Situasjon. Organisasjonen trengte å samle og behandle store mengder tabelldata fra et eksternt datavarehus som lå bak et IPsec-VPN. Disse dataene var avgjørende for interne analysedashbord og eksterne Python-API-er som ble brukt av kunder og partnere.

Oppgave. Målet var å designe, implementere og vedlikeholde et sett med robuste og automatiserte ETL/ELT-pipelines for å hente, transformere og laste data sikkert inn i Google BigQuery, med nøyaktige data, god ytelse og skalerbarhet på tvers av systemene.

Handling. 6 ende-til-ende ETL/ELT-pipelines ble designet, implementert og administrert, på tvers av flere teknologier, blant annet Google BigQuery, MSSQL, PostgreSQL, Shell-skripting, PL/pgSQL og Transact-SQL.

- Etablerte sikre tilkoblinger til en ekstern server beskyttet av et IPsec-VPN for å automatisere datainnhentingen.
- Planla og kjørte nedlastinger av komprimerte dataarkiver med Parquet-, CSV- og .bak-filer.
- Utviklet Bash- og Python-skript for å pakke ut og klassifisere filer etter type og skjema.
- For .bak-filer ble sikkerhetskopiene gjenopprettet i en lokal MSSQL Server-instans med RESTORE DATABASE-arbeidsflyter, og skjemaintegriteten ble validert.

- Lastet tabelldata inn i staging-skjemaer i MSSQL og PostgreSQL med verktøyene bcp, psql og SSIS, avhengig av kildeformatet.
- Skrev modulære og gjenbrukbare PL/pgSQL- og T-SQL-prosedyrer for å rense, normalisere og berike dataene ut fra forretningslogikken.
- Eksporterte kuraterte datasett og tabeller til mellomformater, komprimerte dem med gzip og overførte dem sikkert til en Google Cloud Storage-bucket.
- Automatiserte opplastingen og skjemamappingen til Google BigQuery med bq CLI og Python-baserte skript for datainnlasting.

Resultat. Disse automatiserte pipelinene reduserte det manuelle merarbeidet og behandlingstiden betydelig — fra 3+ timer manuelt arbeid til under 20 minutter fra ende til ende, samtidig som dataenes ferskhet ble forbedret fra ukentlige til daglige synkroniseringer. Python-API-ene som brukte disse dataene, fikk 30 % bedre ytelse, og den bedre oversikten hjalp forretningsanalytikerne med å levere raskere innsikt til interessentene. Løsningen er fortsatt skalerbar og utvidbar for onboarding av nye datakilder etter hvert som virksomheten vokser.

Bruk erfaringen i prosjektet deres

Bruk «Designet og driftet 6 ETL/ELT-pipelines» som utgangspunkt for en lignende utfordring i organisasjonen deres. Tjenestene bak arbeidet er tilgjengelige for prosjektet deres. En struktur som samsvarer med hvordan virksomheten faktisk fungerer, slik at spørsmålene du trenger svar på senere, fortsatt er mulige å stille.

Se tjenestene våre Ta kontakt →

Se bak resultatet

Les om situasjonen, beslutningene og resultatet i eksemplet. Vi kan drøfte hva som passer deres rammer, og hva som krever en annen tilnærming. Designet og driftet 6 ETL/ELT-pipelines — 6 ETL-pipelines på tvers av tre databasemotorer

Se tjenestene våre Ta kontakt →

Trenger du å samle data fra mange kilder?

Snakk om prosjektet med ingeniøren som har ansvaret for leveransen. Vi avtaler omfang, prisgrunnlag og overlevering skriftlig før arbeidet begynner.

Rolle i oppdraget: Dataingeniør. Det bygger på Databaser, PostgreSQL, SQL, Datavarehus, Data engineering, Datapipelines (ETL/ELT), Skytjenester, Automatisering og CI/CD, Nettverk og VPN og Python. Og nå som dere er her: vi er ganske gode på Backend- og API-utvikling, Databasedesign og datamodellering, Utvikling av datapipelines (ETL/ELT), Design av datavarehus og Klargjøring av data for AI.

Ta kontakt Se tjenestene våre

Ett stykke av **en lengre oversikt** — og den ligger i sin helhet her.

Hver oppføring er skrevet på samme måte: situasjonen, oppgaven, hva vi gjorde, og hva som endret seg.

Se hele oversikten

Definer forbedringen dere trenger

Ta med et representativt datasett, spørsmålene det skal besvare og dagens flaskehals. Vi kan avtale målinger av kvalitet, gjennomstrømning og gjenoppretting.

Se hva det koster Ta kontakt →

Trenger du å samle data fra mange kilder?

Spør om beslutningene bak «Designet og driftet 6 ETL/ELT-pipelines». Dere snakker med folk som kan forklare arbeidet og vurdere deres eget prosjekt.

Se tjenestene våre

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

Ta med deg denne siden

Visittkort

Laget med kjærlighet av D.

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

Opphavsrett © 2025 – nå · Engineer ApS · Alle rettigheter forbeholdt.

Skrevet ut fra https://data.engineer.company/nb/portfolio/designed-and-ran-6-etl-elt-pipelines/