Engineer Data

Utformade och drev 6 ETL/ELT-pipelines

Utformade, implementerade och administrerade 6 ETL/ELT-pipelines med Google BigQuery, MSSQL, PostgreSQL, Shell-skript, PL/pgSQL och Transact-SQL och integrerade data för effektiv bearbetning i ett Python-API.

2022 – 2024

6

ETL-pipelines över tre databasmotorer

Situation. Organisationen behövde konsolidera och bearbeta stora volymer tabelldata från ett fjärrbeläget datalager bakom ett IPsec-VPN. Dessa data var avgörande för att driva interna analysdashboards och externa Python-API:er som användes av kunder och partner.

Uppgift. Målet var att utforma, implementera och underhålla en uppsättning robusta och automatiserade ETL/ELT-pipelines för att på ett säkert sätt hämta, transformera och läsa in data i Google BigQuery, och samtidigt säkerställa korrekta data, prestanda och skalbarhet i alla system.

Handling. 6 kompletta ETL/ELT-pipelines utformades, implementerades och administrerades, med en rad tekniker som Google BigQuery, MSSQL, PostgreSQL, Shell-skript, PL/pgSQL och Transact-SQL.

- Etablerade säkra anslutningar till en fjärrserver skyddad av ett IPsec-VPN för att automatisera datahämtningen.
- Schemalade och körde nedladdningar av komprimerade dataarkiv med Parquet-, CSV- och .bak-filer.
- Utvecklade Bash- och Python-skript för att extrahera och klassificera filer efter typ och schema.
- För .bak-filerna återställdes säkerhetskopiorna till en lokal MSSQL-instans med arbetsflöden baserade på RESTORE DATABASE, och schemaintegriteten validerades.

- Läste in tabelldata i staging-scheman i MSSQL och PostgreSQL med verktygen bcp, psql och SSIS, beroende på källformatet.
- Skrev modulära och återanvändbara PL/pgSQL- och T-SQL-procedurer för att rensa, normalisera och berika data utifrån affärslogiken.
- Exporterade kurerade dataset och tabeller till mellanformat, komprimerade dem med gzip och överförde dem säkert till en bucket i Google Cloud Storage.
- Automatiserade uppladdningen och schemamappningen till Google BigQuery med bq CLI och Python-baserade skript för datainläsning.

Resultat. Dessa automatiserade pipelines minskade avsevärt det manuella merarbetet och bearbetningstiden — från över 3 timmars manuellt arbete till under 20 minuter för hela flödet, samtidigt som datafärskheten förbättrades från veckovis till daglig synkronisering. De Python-API:er som använde dessa data fick 30 % bättre prestanda, och den ökade insynen hjälpte affärsanalytikerna att snabbare leverera insikter till intressenterna. Lösningen är fortfarande skalbar och utbyggbar, så att nya datakällor kan anslutas i takt med att verksamheten växer.

Använd erfarenheten i ert projekt

Använd ”Utformade och drev 6 ETL/ELT-pipelines” som utgångspunkt för en liknande utmaning i er organisation. Tjänsterna bakom arbetet finns tillgängliga för ert projekt. En struktur som speglar hur verksamheten faktiskt fungerar, så att de frågor du behöver svar på senare fortfarande går att ställa.

Se våra tjänster Hör av er →

Se bakom resultatet

Läs om situationen, besluten och resultatet i exemplet. Vi kan diskutera vad som passar era förutsättningar och vad som kräver ett annat angreppssätt. Utformade och drev 6 ETL/ELT-pipelines — 6 ETL-pipelines över tre databasmotorer

Se våra tjänster Hör av er →

Behöver du samla data från många källor?

Prata om ert projekt med ingenjören som ansvarar för leveransen. Vi avtalar omfattning, hur priset beräknas och överlämning skriftligt innan arbetet börjar.

Roll i uppdraget: Dataingenjör. Det bygger på Databaser, PostgreSQL, SQL, Datalager, Data engineering, Datapipelines (ETL/ELT), Molntjänster, Automatisering och CI/CD, Nätverk och VPN och Python. Och nu när ni är här: vi är ganska bra på Backend- och API-utveckling, Databasdesign och datamodellering, Utveckling av datapipelines (ETL/ELT), Design av datalager och Dataförberedelse för AI.

Hör av er Se våra tjänster

En del av **en längre historik** — och den finns i sin helhet här.

Varje post är skriven på samma sätt: situationen, uppgiften, vad vi gjorde och vad som förändrades.

Se hela historiken

Definiera förbättringen ni behöver

Ta med ett representativt dataset, frågorna det ska besvara och dagens flaskhals. Vi kan avtala mått på kvalitet, genomströmning och återställning.

Se vad det kostar Hör av er →

Behöver du samla data från många källor?

Fråga om besluten bakom ”Utformade och drev 6 ETL/ELT-pipelines”. Ni talar med personer som kan förklara arbetet och bedöma ert eget projekt.

Se våra tjänster

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

Ta med dig den här sidan

Visitkort

Omsorgsfullt skapad med kärlek av D.

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

Upphovsrätt © 2025 – nu · Engineer ApS · Alla rättigheter förbehållna.

Utskrivet från https://data.engineer.company/sv/portfolio/designed-and-ran-6-etl-elt-pipelines/