Engineer Дані

Спроєктувала й підтримувала 6 пайплайнів ETL/ELT

Спроєктувала, реалізувала та адмініструвала 6 пайплайнів ETL/ELT, використовуючи Google BigQuery, MSSQL, PostgreSQL, shell-скрипти, PL/pgSQL та Transact-SQL, інтегрувавши дані для ефективної обробки через Python API.

2022 – 2024

6

пайплайнів ETL на трьох рушіях баз даних

Ситуація. Організації потрібно було консолідувати та обробляти великі обсяги табличних даних із віддаленого сховища даних, розташованого за IPsec VPN. Ці дані мали критичне значення для роботи внутрішніх аналітичних дашбордів і зовнішніх Python API, якими користувалися клієнти й партнери.

Завдання. Мета полягала в тому, щоб спроєктувати, реалізувати та підтримувати набір надійних автоматизованих ETL/ELT-пайплайнів, які безпечно отримують, трансформують і завантажують дані в Google BigQuery та забезпечують точність даних, продуктивність і масштабованість у всіх системах.

Дія. Спроєктовано, реалізовано й адміністровано 6 наскрізних ETL/ELT-пайплайнів, що охоплювали кілька технологій, зокрема Google BigQuery, MSSQL, PostgreSQL, Shell-скрипти, PL/pgSQL і Transact-SQL.

- Налагоджено безпечні з’єднання з віддаленим сервером за IPsec VPN, щоб автоматизувати отримання даних.
- Заплановано й виконано завантаження стиснутих архівів даних, що містили файли Parquet, CSV та .bak.
- Розроблено скрипти на Bash і Python для розпакування й класифікації файлів за типом і схемою.

- Резервні копії з файлів .bak відновлено в локальному екземплярі MSSQL Server за допомогою процедур RESTORE DATABASE і перевірено цілісність схеми.
- Завантажено табличні дані в проміжні схеми в MSSQL і PostgreSQL за допомогою інструментів bcp, psql та SSIS, залежно від формату джерела.
- Написано модульні та придатні для повторного використання процедури PL/pgSQL і T-SQL для очищення, нормалізації та збагачення даних відповідно до бізнес-логіки.
- Експортовано впорядковані набори даних і таблиці в проміжні формати, стиснуто їх за допомогою gzip і безпечно передано в бакет Google Cloud Storage.
- Автоматизовано процеси завантаження та зіставлення схем у Google BigQuery за допомогою bq CLI і Python-скриптів для приймання даних.

Результат. Ці автоматизовані пайплайни суттєво скоротили обсяг ручної роботи й час обробки — з понад 3 годин ручної роботи до менш ніж 20 хвилин на весь цикл, а синхронізація даних стала щоденною замість щотижневої. Продуктивність Python API, які використовували ці дані, зросла на 30%, а краща видимість даних допомогла бізнес-аналітикам швидше доносити висновки до зацікавлених сторін. Рішення залишається масштабованим і легко розширюваним для підключення нових джерел даних у міру зростання бізнесу.

Застосуйте цей досвід у своєму проєкті

Використайте «Спроєктувала й підтримувала 6 пайплайнів ETL/ELT» як відправну точку для схожого завдання у вашій організації. Послуги, що стоять за цією роботою, доступні й для вашого проєкту. Структура, що відповідає реальній роботі бізнесу, тож потрібні згодом запитання до даних усе ще можна буде поставити.

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. Спроєктувала й підтримувала 6 пайплайнів ETL/ELT — 6 пайплайнів ETL на трьох рушіях баз даних

Переглянути наші послуги Зв'язатися з нами →

Потрібно об'єднати дані з багатьох джерел?

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 даних. Тут задіяні Бази даних, PostgreSQL, SQL, Сховища даних, Інженерія даних, Пайплайни даних (ETL/ELT), Хмарні технології, Автоматизація та CI/CD, Мережі та VPN та Python. І, до речі, ми добре вміємо Backend- та API-розробка, Проєктування та моделювання баз даних, Розробка пайплайнів даних (ETL/ELT), Проєктування сховищ даних та Підготовка даних для ШІ.

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 →

Потрібно об'єднати дані з багатьох джерел?

Запитайте про рішення в роботі «Спроєктувала й підтримувала 6 пайплайнів ETL/ELT». 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/designed-and-ran-6-etl-elt-pipelines/