Engineer Données

6 pipelines ETL/ELT conçus et exploités

Conçu, mis en œuvre et administré 6 pipelines ETL/ELT avec Google BigQuery, MSSQL, PostgreSQL, des scripts Shell, PL/pgSQL et Transact-SQL, en intégrant les données pour leur traitement efficace par une API Python.

2022 – 2024

6

pipelines ETL sur trois moteurs de bases de données

Situation. L'organisation devait réunir et traiter de grandes quantités de données tabulaires provenant d'un entrepôt distant accessible derrière un VPN IPsec. Ces données étaient essentielles aux tableaux de bord analytiques internes et aux API Python externes utilisées par les clients et partenaires.

Tâche. L'objectif était de concevoir, mettre en œuvre et maintenir des pipelines ETL/ELT solides et automatisés pour récupérer, transformer et charger les données dans Google BigQuery en sécurité, avec une exigence de précision, de performances et de montée en charge entre systèmes.

Action. 6 pipelines ETL/ELT de bout en bout ont été conçus, mis en œuvre et administrés avec Google BigQuery, MSSQL, PostgreSQL, des scripts Shell, PL/pgSQL et Transact-SQL.

- Des connexions sécurisées à un serveur distant protégé par VPN IPsec ont été établies pour automatiser la récupération des données.
- Les téléchargements d'archives compressées contenant des fichiers Parquet, CSV et .bak ont été planifiés et exécutés.
- Des scripts Bash et Python ont été développés pour extraire et classer les fichiers selon leur type et leur schéma.
- Les sauvegardes .bak ont été restaurées dans une instance MSSQL Server locale avec RESTORE DATABASE, puis l'intégrité du schéma a été validée.
- Les données tabulaires ont été chargées dans des schémas de préparation MSSQL et PostgreSQL avec bcp, psql et SSIS, selon le format source.

- Des procédures PL/pgSQL et T-SQL modulaires et réutilisables ont été écrites pour nettoyer, normaliser et enrichir les données selon la logique métier.
- Les jeux et tables préparés ont été exportés dans des formats intermédiaires, compressés avec gzip et transférés de façon sécurisée vers un bucket Google Cloud Storage.
- Les transferts vers Google BigQuery et la correspondance des schémas ont été automatisés avec bq CLI et des scripts d'ingestion Python.

Résultat. Ces pipelines ont fortement réduit le travail manuel et la durée des traitements : 3 heures ou plus de travail manuel sont devenues moins de 20 minutes de bout en bout, tandis que les synchronisations sont passées d'hebdomadaires à quotidiennes. Les API Python consommant ces données ont gagné 30 % de performances et la meilleure visibilité a permis aux analystes de fournir plus vite leurs analyses aux parties prenantes. La solution reste évolutive et extensible pour intégrer de nouvelles sources avec la croissance de l'entreprise.

Mettez cette expérience au service de votre projet

Prenez « 6 pipelines ETL/ELT conçus et exploités » comme point de départ pour un défi similaire dans votre organisation. Les prestations mobilisées pour cette réalisation sont disponibles pour votre projet. Une structure fidèle au fonctionnement de l'entreprise, qui permet encore de poser demain les questions dont vous aurez besoin.

Découvrir nos prestations Contactez-nous →

Examinez ce qui a conduit au résultat

Découvrez la situation, les décisions et le résultat de cet exemple. Nous pouvons discuter des éléments transposables à vos contraintes et de ceux qui nécessitent une autre démarche. 6 pipelines ETL/ELT conçus et exploités — 6 pipelines ETL sur trois moteurs de bases de données

Découvrir nos prestations Contactez-nous →

Des données de nombreuses sources à réunir ?

Discutez de votre projet avec l'ingénieure responsable de sa réalisation. Nous convenons par écrit du périmètre, de la base tarifaire et de la transmission avant le début des travaux.

Rôle exercé dans ce contexte : Ingénieure des données. Cette réalisation fait appel à Bases de données, PostgreSQL, SQL, Entrepôts de données, Ingénierie des données, Pipelines de données (ETL/ELT), Cloud, Automatisation et CI/CD, Réseaux et VPN et Python. Et puisque vous êtes ici, nous sommes aussi plutôt doués en Développement backend et API, Conception et modélisation de bases de données, Développement de pipelines de données (ETL/ELT), Conception d'entrepôts de données et Préparation des données à l'IA.

Contactez-nous Découvrir nos prestations

Un élément **d'un parcours plus long** — entièrement présenté sur ce site.

Chaque entrée suit la même structure : la situation, la tâche, ce que nous avons fait et ce qui a changé.

Voir l'ensemble du parcours

Définissez l'amélioration nécessaire

Présentez un jeu de données représentatif, les questions auxquelles il doit répondre et le goulot d'étranglement actuel. Nous pouvons convenir de la manière de mesurer la qualité, le débit et la reprise.

Consulter les tarifs Contactez-nous →

Des données de nombreuses sources à réunir ?

Interrogez-nous sur les décisions qui sous-tendent « 6 pipelines ETL/ELT conçus et exploités ». Vous parlerez aux personnes capables d'expliquer le travail et d'évaluer votre propre projet.

Découvrir nos prestations

welcome@engineer.company +45 50 10 61 84 Copenhague, Danemark

Emportez cette page avec vous

Fiche de contact

Créé avec soin et amour par D.

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

Droit d'auteur © 2025 – aujourd'hui · Engineer ApS · Tous droits réservés.

Imprimé depuis https://data.engineer.company/fr/portfolio/designed-and-ran-6-etl-elt-pipelines/