Database replication

Health & wellness firm

Replicated the existing database of the client's subsidiary company in an Azure-based Data Warehouse, integrating it into their central Data Warehouse with visibility over the brand data, thus providing centralized reporting

Health & wellness firm needs database replication

Picture this…

You're looking to replicate the existing database in an Azure-based Database within the company's central Data Warehouse to provide the client with complete visibility over the brand data, allowing them to use it for central reporting. Acquired a brand whose existing Data Warehouse was outside their ecosystem, hindering central reporting at the group level

You turn to Accordion.

We partner with your team to Replicated the existing database of the client's subsidiary company in an Azure-based Data Warehouse, integrating it into their central Data Warehouse with visibility over the brand data, thus providing centralized reporting, including:

1) Replicating the Azure-based Data Warehouse and implementing an optimized daily data refresh process to selectively update tables that were modified at the source, thereby enhancing efficiency in data synchronization
2) Developing and implementing an automated alert system to instantly notify the team upon the creation of a new source table in the child company's warehouse, providing relevant details.
3) Incorporating automated data validation and reconciliation rules to ensure accuracy and consistency in the data load process

Your value is enhanced.

- Ensured consistent data across both data warehouses with 100% accuracy during the migration process, guaranteeing the reliability of the dashboarding suite at the client's end as a source of truth. Achieved efficient daily data refresh by only updating the data tables modified since the last refresh
- Automated data refresh saved ~90 hours of person-hours by eliminating the need for manual data refresh every week

DATABASE REPLICATION

KEY RESULT

- Saved ~90 hours of person-hours per week

VALUE LEVERS PULLED

- One-Time full database copy
- Daily incremental data refreshes

Database replication for a home & personal care firm

Situation

- The client acquired a brand whose existing Data Warehouse was outside their ecosystem, hindering central reporting at the group level
- Partnered with the client to replicate the existing database in an Azure-based Database within the company's central Data Warehouse to provide the client with complete visibility over the brand data, allowing them to use it for central reporting

Accordion Value Add

- Replicated the Azure-based Data Warehouse and implemented an optimized daily data refresh process to selectively update tables that were modified at the source, thereby enhancing efficiency in data synchronization
- Developed and implemented an automated alert system to instantly notify the team upon the creation of a new source table in the child company's warehouse, providing relevant details.
- Incorporated automated data validation and reconciliation rules to ensure accuracy and consistency in the data load process

Impact

- Ensured consistent data across both data warehouses with 100% accuracy during the migration process, guaranteeing the reliability of the dashboarding suite at the client's end as a source of truth. Achieved efficient daily data refresh by only updating the data tables modified since the last refresh
- Automated data refresh saved ~90 hours of person-hours by eliminating the need for manual data refresh every week

Methodology / Approach

01 **Identified** a list of 500+ **tables for replication** based on client's reporting requirements

02 **Designed a loading strategy** by categorizing tables into full load or incremental for daily data refresh

03 For full load tables, performed daily truncation and data loading. For incremental tables, further categorized them into Change Tracking (CT) tables and tables with updated-at columns

04 CT Tables stores information regarding data changes (inserts, updates, deletes) in the main table and we used this information to refresh data from the source and for Tables which has an updated date column, we updated the data since the last refresh based on the updated date column.

05 Created a mapping table using Metadata Tables to mark tables that were updated since the last refresh

06 **Parameterized Azure Data Factory (ADF) pipelines to optimize the daily refresh process** using information from the mapping table

07 Established an **Audit table to ensure the completion of the data load process**, verifying row counts for each table against the source

Data warehouse architecture - High level diagram