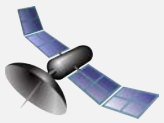
gluent.

Dataset Deduplication: Telco Eliminates Enterprise Data Sprawl with Gluent Data Platform



INDUSTRY

Telecommunications

CHALLENGE

- Dataset sharing required to 20+ databases across the enterprise
- Complex ETL and replication processes introduce potential for decisions made using "stale" data
- Storage and licensing costs continue to rise as more departments request duplicates of the data

SOLUTION

- Implement Hadoop cluster as a centralized data store
- Share datasets by offloading to Hadoop via Gluent Offload
- Transparently present offloaded data to 20+ databases with no ETL

RESULTS

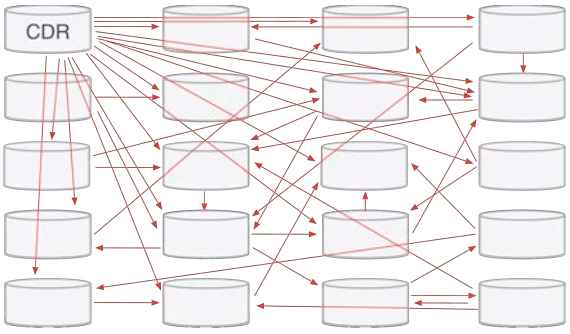
- Eliminated data sprawl and duplication of datasets
- Plan to reduce licensed RDBMS by half
- Dozens of complex ETL and data replication processes removed

Enterprise data rarely remains only in its database of origin these days. Different business divisions or departments require the data for further reporting, analysis, decision making, or simply to drive business rules in other applications. In large enterprises, with data centers spread throughout the world, access to the original source data, even if the "source" happens to be an enterprise data warehouse, may not be simple to implement. This is where enterprise data sprawl often begins, with copies of the data being duplicated in multiple databases throughout the organization.

ENTERPRISE DATA SPRAWL

A Gluent customer, a global telecommunications company, happened to be deep into the data sprawl dilemma. In the telecom industry, the primary dataset is the call data record (CDR). The CDR contains information about calls and text messages between telephones and is used for many different purposes throughout the organization. This dataset was copied from its original source to 20+ other relational databases throughout the enterprise. This is less than ideal for several reasons:

- Duplication of data to many databases requires some sort of ETL or replication process and a lot of moving parts. If at any point one or more of these data movement mechanisms fails, there is a potential for discrepancies in the "freshness" of the data across the enterprise. This could lead to some decisions being made using out-of-date information.
- Storage space must be considered when storing multiple copies of the same data. And not just the storage cost, but also the database licenses required to run 20+ databases.
- Data shared from the CDR dataset to many additional databases, and data from those schemas shared further around, created the complex spider web of ETL processes you see in the below image.



As a side note, the dataset was rapidly growing and continued to put a strain on many of the database servers when certain advanced analytics or historical queries were executed. Database servers often couldn't keep up with the demands of the business questions being asked.

The customer had quite an intricate data sharing setup to begin with, and Gluent Data Platform was able to simplify it tremendously. This is a common pattern we see across major enterprises, most notably at large telcos such as this one. Gluent Data Platform is a data sharing and data virtualization platform, designed to centralize all enterprise data while continuing to allow virtual data access from the database of origin, without any changes to existing applications or reports.

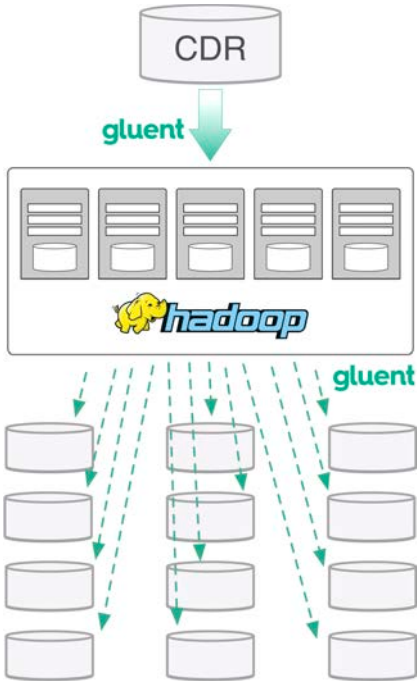
DATA SHARING AND DATA VIRTUALIZATION

Two major functions of Gluent Data Platform were implemented to reduce the spread of dataset copies throughout the enterprise:

- **Data offload**, via Gluent Offload Engine, copying data from the RDBMS to Hadoop, a modern, distributed storage and compute engine.
- **Data access**, using Gluent Present to share any table that exists in Hadoop with any relational database.

Using these two capabilities, Gluent Data Platform could provide data sharing and data virtualization across the enterprise. Data virtualization has many different meanings (as you can see on Wikipedia), depending on the goal of the virtualization software. In the case of Gluent, we consider our platform to "truly" virtualize data. With Gluent, you can transparently plug the power of Hadoop into your existing application without any upfront development investment. No application code changes or ETL projects required. Therefore, you can determine whether Gluent works for your application (and business) by simply installing and configuring Gluent software... as opposed to a large upfront investment, replatforming your application and rewriting code, just to see if the new platform works for you. Gluent Data Platform can, and should, be tried out in days and weeks, not months and years. With that, let's get back to the challenge at hand.

Gluent Data Platform will allow all data to be centrally stored in Hadoop and, if desired, dropped from source database. In many other customer use cases, dropping the data from the original database is the best way to eliminate storage, CPU, and even license cost for the RDBMS. But in the case of our large telco customer, they simply took advantage of the data sharing capabilities of Gluent. Data sharing requires zero additional copies of the data after it has been initially offloaded. Using the Present command, application databases were able to access the data stored in Hadoop direct from the Oracle database schemas.



DATASET DEDUPLICATION RESULTS

The customer could access the call data record dataset from any database via Gluent Present, and the copies throughout the enterprise could be removed completely. Storage and CPU savings led to smaller databases - and have a plan in place to eventually reduce the number of active production databases down to nearly half of the original 20+ that were serving up copies of the CDR dataset. This allows for the reallocation of database licenses and will lead to even further cost savings.

Data sharing, plus true data virtualization, provides Gluent customers with a simple solution to eliminate data sprawl throughout their enterprise. For our large telecommunications customer, the ease with which they could "distribute" the centralized dataset was key in reducing costs across the enterprise.

gluent.

World HQ
1701 North Market, #330
Dallas, Texas 75202
United States

+1 (469) 619-7052
info@gluent.com
gluent.com

Gluent is a registered trademark of Gluent Inc. All other trademarks or service marks are the property of their respective holders and are hereby acknowledged. ©2017 Gluent Inc. All rights reserved.