MIGRATING LEGACY TERADATA DATA WAREHOUSE TO CLOUD

Executive Summary
Enterprise Data Warehouses (EDWs) are central repositories of integrated data purpose-built and optimized to support reporting and analytics. EDWs store current and historical data in an integrated repository to enable business intelligence throughout the enterprise. Building these warehouses have traditionally been multi-year projects, with great thought and data management best practices leveraged to build them.

Many of these EDWs currently run on Teradata’s database platform—consistently rated as one of the top providers of DW systems by Gartner. Our clients have developed these Teradata-based EDWs during the past few decades and many of these platforms store multiple PBs of data and support mission critical reporting, analytic, and operational applications in the enterprise. Growth in data volume, higher cost of running on-premise environments, increasing need to process unstructured data and more extensive use of data science and AI-ML puts a stress on these Teradata based EDW platforms. At the same time, cloud computing has become the de facto platform for fueling digital transformations and modernizing IT portfolios.

Our clients are now considering and planning to migrate these Legacy Teradata EDWs to the cloud—both to alleviate the pain points of the current EDW environments, as well as to take advantage of cloud and its ease of use, agility, flexibility, and often cost savings.

We will refer to these configurations as Legacy Teradata EDWs (to differentiate from the new generation of Teradata Vantage solution). In existence for 30-plus years, Teradata has an extremely mature feature set that caters to the most demanding requirements of some of the largest organizations in the world. The value of its database solution can only be realized at scale.
WHY MIGRATE
The digital revolution currently under way is driving a huge growth in global data volumes. By 2025, IDC estimates worldwide data will grow 61 percent to 175 zettabytes, with as much of the data residing in the cloud as in data centers. 70 percent to 80 percent of the data being generated today is unstructured and doubling every two years. At the same time, there is a growing need for advanced analytics—artificial intelligence, machine learning, and other cognitive services, which are essential for companies to become intelligent enterprises. Legacy Teradata, while an excellent and mature platform, was built on keeping relational data in mind.

Challenges with the current Legacy Teradata environment and a growing need to leverage cloud has created the perfect tsunami that is driving a migration of these legacy Teradata environments to cloud.

**Performance and Capacity Issues**

Teradata is an MPP system that delivers linear scalability. The current technology and architecture of Teradata (not including Vantage) is being considered “legacy.”

**New Data Sources and Workloads**

There is an increasing need to process unstructured data—log files, genomic data, health records, call center notes, documents, and images to name a few. Legacy Teradata is a relational DBMS and not architected and built to handle varying data types and real-time data.

**High Licensing Cost and TCO**

Teradata is the biggest and longest-established player in the data warehouse appliance market. Market research indicates the cost of managing a TB in Legacy Teradata can range from $30,000-$90,000, considerably high compared to today’s cloud pricing.
The Challenge
A large financial institution in North America decided to move all of its applications to the cloud and close out its data centers. As part of this move, a large scale Teradata data warehouse environment was successfully migrated to the cloud.

Case study

The Solution
Technologies in play: Teradata, DB2, Ab Initio, SAS, S3, Snowflake, Redshift, PostgreSQL
- 18 PB data migrated
- 15,000 Teradata tables migrated
- 26,000 Ab Initio ETL code graphs re-engineered
- 35,000 ETL scheduler jobs migrated

The Outcome
• 2/3rd reduction in critical incident resolution time
• From 1 day to 2 hrs for disaster recovery
• From 3 months to 3 days to build and release features to production
• 2/3rd reduction in cost per petabyte of storage
• 50% reduction in cost for compute

Movement to Cloud
More and more clients are moving to the public cloud. They want to consume infrastructure and platform as “services” and in an on-demand, pay-as-you-go model. Nearly every enterprise that uses Legacy Teradata today is exploring options to re-platform away from the current data warehouse platform.
CHALLENGES OF A LEGACY TERADATA MIGRATION
Migrating from an on-premises Legacy Teradata EDW to something equivalent in the cloud presents substantial challenges.

Incomplete Knowledge and Understanding

Teradata EDWs have been developed and enhanced during the past 10-20 years. Its systems are used in demanding multi-purpose, multi-tenant scenarios. Planning a migration without a detailed understanding of the current environment is a daunting task.

Complexity of the Environment

Legacy Teradata environments are usually extremely large (from 100s of TBs to multiple PBs) complex and support mission critical analytic applications. Data models can be extremely complex given their customization during the past few decades, and it is not uncommon to see tens of thousands of tables.

Consumption Workloads

Teradata platforms integrate data from multiple data sources and provide a consolidated view of data. Numerous consumption workloads via tools such as reporting/BI (Cognos, Business Objects), visualization (Tableau, Qlik), and analytics (SAS) have been integrated into Teradata.

Lack of Similar Capability in Target Platforms

Not all data warehouse platforms are created equal. Even though multiple MPP cloud data warehouse platforms exist, they may or may not have features and functions like Teradata.

Need to Modernize while Migrating

Clients are often thinking of migrating out of Legacy Teradata as an opportunity to modernize their EDW environment. A choice that clients will face will be to either simply Translate or Transform.
ELEMENTS OF
A MIGRATION
PROJECT
Migrating the Legacy Teradata EDW to another platform in the cloud is a complex, multi-stage process, requiring consideration and planning. Below are elements of a typical migration project.

**Schema**
Schema conversion from Teradata to target, data types do not match 1:1

**Users & Access Control**
Users and access control on various database objects need to be applied

**Stored Procs & BTEQ Scripts**
Convert Teradata stored procs to target SQL and BTEQ commands to SQL, Python, Java or script

**Testing & Validation**
Testing of target platform for data accuracy and validation of results from BI reports and other consumption

**Capacity Planning**
Capacity needed in target (nodes, processors, memory, disk storage) to meet SLA

**Initial Data**
Bulk movement of historical data before cutover - extract/load, appliance-based transfer or physical movement based on volume

**Ingestion & ETL Pipelines**
Data ingestion pipelines from different sources and subsequent ETL/ELT code migration

**Queries & BI Workloads**
SQL query migration accounting for Teradata specifics, re-pointing of BI tools and any changes to existing reports and dashboards
We have established the reasons for migrating away from Legacy Teradata and what it involves, but there are a few options for target platforms, including:

**Option 1**  
**Teradata Vantage in Cloud**

Teradata Vantage is an analytical platform that combines open source and proprietary services to offer a holistic analytical capability. Teradata, while delayed reacting to the cloud, understands the changing data landscape.

**Option 2**  
**Cloud Native**

The market for cloud data warehouses has grown in recent years—both in terms of size and maturity, as organizations move to take advantage of cloud economics and reduce their own physical data center footprints. Each of the three major cloud service providers now has a mature data warehouse platform offering—Amazon AWS with Redshift, Microsoft Azure with Synapse Analytics, and Google Cloud with BigQuery.

**Option 2**  
**Third-Party PaaS**

There are other options for clients looking to migrate to the cloud but don’t want to commit to cloud-native data warehouse services. A key reason for this can be a need for cloud independence and portability across clouds. There is no shortage of choices for a cloud-based data warehouse. As part of this option there are five major choices: Snowflake Cloud Data Warehouse, Cloudera Data Platform (CDP), Databricks, Yellowbrick, and IBM Cloud Pak for Data. All these platforms are available on public cloud vendors (AWS, Azure, and Google).

The decision on which cloud provider to select is mostly a strategic one, and generally not driven by the data warehouse. **It will depend on a specific client scenario, drivers for migration, and the other criteria.**

Migration of a Legacy Teradata Data Warehouse from on-premise to the cloud is not for the faint-hearted. As highlighted, there are a lot of complexity and hidden challenges. Depending on the size and scale of the environment, the target platform chosen, and amount of re-engineering to be performed, the migration could be a multi-year journey.
Plan Your Migration  Execute in Phases  Manage Change

A data warehouse is a critical component of any enterprise’s data and analytics landscape and a number of these large-scale data warehouses have been developed on-premise using Teradata Data Warehouse platforms. These platforms have served the reporting and analytics needs during the past 10-20 years and have grown to store PBs of data.

Factors such as a higher TCO of current environment, need to adopt cloud computing, exploding data volumes, and requirements to support AI and Machine Learning at scale, are all driving enterprises to consider migrating and modernizing their Legacy Teradata data warehouses to cloud. Migration approaches can range from simple translate to full transformation, and this will drive the complexity, timeline and benefits. Careful upfront planning, phased migration execution, and use of automation tools can alleviate the risk and improve chances of success.
ABOUT ACCENTURE

Accenture is a leading global professional services company, providing a broad range of services in strategy and consulting, interactive, technology and operations, with digital capabilities across all of these services. We combine unmatched experience and specialized capabilities across more than 40 industries — powered by the world’s largest network of Advanced Technology and Intelligent Operations centers. With 513,000 people serving clients in more than 120 countries, Accenture brings continuous innovation to help clients improve their performance and create lasting value across their enterprises. Visit us at www.accenture.com.