



## Treehouse Dataflow Toolkit: Rapid Mainframe Data Transfer to Analytics/ML/AI-Friendly Targets on AWS via Amazon MSK

### About the Treehouse Dataflow Toolkit...

Treehouse Dataflow Toolkit (TDT) is a set of proprietary microservices that assure highly-available, auto-scalable, and event-driven data transfers to your data science teams' favorite analytics frameworks.

Customers either already have, or are in the process of acquiring, software tools that replicate their data into Kafka pipelines. TDT provides the turn-key solution for getting this data from Kafka into advanced Analytics/AI/ML-friendly targets, such as Amazon Redshift, Snowflake, Amazon Athena/S3, Amazon S3 Express One Zone Buckets, as well as Amazon Aurora PostgreSQL, all the while **adhering to AWS's and Snowflake's recommended best practices for massive data loading**, thus assuring shortest and surest loads.

TDT not only keeps things up to date faster than any conceivable ODBC-based solution, but the "delta tables" into which it loads data also inherently retain the entire history of source data ever since mainframe-to-target synchronization began. So, for example, after TDT has been syncing a target table for 5 years, **a data scientist now has 5 years' worth of historical data to work with for trend analysis, predictive analytics, prescriptive analytics, ML, etc.**



### Benefits

#### Performance at scale

TDT provides massive scalability and reliability, thanks to the AWS Lambda infrastructure.

#### High-speed throughput

A delta-table approach means unbeatable throughput (everything is an INSERT, and is going through the target vendors' "best-practices" bulk-load utilities).

#### Advanced crawler (DDL generator) technology

TDT's proprietary crawler functions are used to automatically prepare landing tables, views, and additional ELT infrastructure for various analytics-friendly targets. While data is landed in targets in JSON format (a natural fit for hierarchically-structured data from many mainframe sources), a full set of user-views is generated for those who prefer to manipulate the data in traditional, fully-structured relational format.

#### Cost effective

AWS solutions provide the security, availability, and reliability of data stores at one-tenth of the cost. You only pay for the compute resources used during the migration process and any additional log storage.

## AWS data store solutions make it easy

Companies, government agencies, educational institutions, and nonprofit organizations are relying on Amazon Web Services (AWS) to quickly and securely accelerate the migration of databases. By using AWS solutions, tools, programs, and databases available, you can save time and automate processes. AWS provides performance at scale, reliable databases, cost-effective solutions, and fully managed services that give you the flexibility when it comes to administration.



## Explore the portfolio of AWS database solutions and benefits

Amazon Web Services (AWS) database solutions ease the migration process and its complexities for companies, government agencies, educational institutions, and nonprofit organizations.

AWS Database Migration Service (AWS DMS) supports both homogeneous, and heterogeneous migrations between different database platforms, while the AWS Schema Conversion Tool (AWS SCT) helps to predict heterogeneous database migrations.

Other solutions like the Amazon Database Migration Accelerator (Amazon DMA) bring together experts, mechanisms, and tools to provide a one-stop shop to convert applications, while the Database Freedom program helps to cut down the heavy lifting that is involved in database administration. These four resources make up the AWS portfolio of fully managed, high performance, and cost-effective solutions for database migration.

## Treehouse Software Resources

### [TDT Blog](#)

Treehouse Software blog gives a high-level overview of TDT, as well as discusses the market and competitive differentiators.

### [Mainframe-to-AWS Data Replication Tools from Treehouse Software](#)

Treehouse Software offers the most comprehensive and flexible portfolio of solutions available anywhere for integration and replication of data between mainframe sources and any target, application or platform using ETL, CDC, SQL, etc. technologies.



## Sign up for a low-risk Treehouse Software POC...

Discover how quickly you can begin replicating your mainframe data on AWS during a Proof of Concept (POC). A POC is approximately 5-10 business days, with the customer providing use case and goals for the POC. A Treehouse Software consultant will assist in conducting a limited-scope implementation. This application uses a sample subset of customer data and executes in a non-production environment. A document is provided beforehand that outlines the requirements and agenda for the POC. By the end of the 10-day POC, customers can begin replicating mainframe data to their AWS target data store. It can happen that fast!

Contact Treehouse Software Today

