

TIMEXTENDER

touchstone  
BUSINESS INTELLIGENCE



# HOW TO BUILD A FUTURE-PROOF DATA LAKE 10X FASTER

---

Achieve Automated, Low-Code Data  
Integration with TimeXtender



# HOW TO BUILD A **FUTURE-PROOF** DATA LAKE 10X FASTER

## Achieve Automated, Low-Code Data Integration with TimeXtender

A data lake is a type of data storage that allows for the accumulation of large amounts of data in its native, raw format. By storing data in its raw, unstructured form, a data lake gives data scientists, engineers, analysts, and developers the ability to quickly and easily access all of an organization's data, regardless of its source or format.

This flexibility makes data lakes particularly well suited for exploring Big Data sets. In addition, data lakes can be used to support a variety of data-driven applications, such as data mining, machine learning, and predictive analytics. As a result, data lakes can provide significant benefits for organizations that are looking to make the most of their data.

A data lake can be built on-premises, in the cloud, or hybrid. It is a scalable solution that can be easily adjusted to accommodate changing data needs.



# TOP 5 BENEFITS OF A DATA LAKE

Data lakes offer many benefits, including:

1

**Increased Efficiency and Productivity:** Data lakes make it easier for organizations to access data and put it to use. In the past, data was often siloed in different departments or data warehouses, making it difficult to get a holistic view of the business. With a data lake, all data is centralized and can be easily accessed by anyone in the organization.

2

**Cost Savings:** Data lakes can help organizations save money by avoiding the need to purchase expensive data warehouse solutions. In addition, data lakes are usually deployed on commodity hardware, which further reduces costs.

3

**Improved decision making:** With all data in one place, organizations can make better-informed decisions. Data lakes make it possible to quickly analyze large data sets and identify trends that would otherwise be difficult to spot.

4

**Increased agility:** Data lakes improve organizational agility by allowing businesses to quickly adapt to changing market conditions. They also make it easier to experiment with new data-driven initiatives without having to go through a lengthy approval process.

5

**Greater insights:** Data lakes make it possible to gain insights that would otherwise be hidden in siloed data sets. By bringing together data from different sources, businesses can generate new insights that can be used to improve products and services or create new revenue streams.





# A TECHNICAL NIGHTMARE

While data lakes offer many benefits, they also have a number of downsides. One such downside is the need for skilled data engineers to hand-code data pipelines in order to extract insights from the data. Today's modern data lake is a loose concept, so each lake is built differently. And to build one, your data team will need to master a wide variety of technologies.

Does your team have skills in R, Python, Hive, NoSQL, or Parquet? How about Hadoop, Sqoop, Pig, Kafka, Scala, or Avro? This is just a small list of the technology skillsets that you would either need to hire or acquire to even start planning out a data lake.

But, let's say you do skill up your team, hire a few PhDs and some consultants and someone build this structure. Then comes the next phase, maintenance. If the data changes or needs to be updated, the data engineer may need to go back and make changes to the pipeline, which can be difficult and time-consuming, especially if custom code has been written, but the original developer is no longer available.

***“A homegrown IT approach may bring an initial 20 percent cost reduction, but result in a 200 percent increase in maintenance costs.” – Gartner***

## FACT: DATA SOURCES WILL KEEP RENEWING, GROWING, AND CHANGING

So, you need to consider not only your current data but the speed at which your data is growing and/or changing. Here are some examples of what will most probably happen with data in your data sources:

- New tables / fields being added
- Tables / fields being renamed
- Tables / fields being deleted
- Data type / structure changes
- Data sources being changed / updated
- New data sources being added

It is obvious this heavily taxes your scarce and expensive resources when they constantly need to hardcode & re-engineer fragile pipelines, adjust API calls, and update connectors. This makes it almost impossible for data engineers to keep up with this if you choose to do it manually.



# DO YOU WANT TECHNOLOGY OR A SOLUTION?

Before you start building your data lake you must ask yourself if you want technology or a solution. Building the data lake is not about “this approach versus that approach” or “what technology is better”. It’s about identifying what business problem you are trying to solve with the data lake, and then delivering the solution for that as fast as possible, and in the most future-proof way possible. Part of prioritizing progress over perfection is using a methodology that encourages this sort of behavior. The first consideration should be whether you want to continuously reinvent the wheel yourself or whether you want automation to work for you.



## AUTOMATION IS THE KEY

As we’ve established, data lakes are traditionally built through a process of hand-coding, which can be time-consuming and expensive. However, data lakes can now be built using automated data integration tools. An automated data integration tool will allow you to quickly and easily connect to any data source, without the need for hand-coding. This means that data engineers can focus on more important tasks, such as data modeling and data analysis.

In addition, data integration tools like TimeXtender can automatically detect changes in data sources and make the necessary changes to data pipelines, without the need for human intervention. This means that data lakes can be built 10x faster and are more future-proof, as they can easily adapt to changes in data sources.





# HOW **TIMEXTENDER** ENABLES DATA TEAMS TO BUILD FUTURE-PROOF DATA LAKES 10X FASTER

TimeXtender solves the problems mentioned above using automation. The TimeXtender Operational Data Exchange (ODX) can **automatically synchronize** with your data sources. TimeXtender will automatically synchronize the structure of the source with the metadata stored in the ODX repository.

## Once TimeXtender detects a change in your data structure it will automatically:

- Create a new version based on the new and/or changed structure of the specific table in the data lake
- Initiate a full load for the specific table to backfill the data if applicable
- Initiate a full load for the specific table in the data warehouse to propagate the change integral in your data estate
- Automatically switch back to incremental load, if available
- TimeXtender automatically uses the latest version of your data

And there is more...

- The TimeXtender ODX Server creates **multiple versions** of your source data. This can be triggered by a detected change in the data structure or by the option to store every scheduled transfer as a new version. This provides a great benefit for the business, because now they'll automatically have their backup and history created which may be used for recovery

and or making older versions of the data (structure) available.

- With the automation in TimeXtender it is possible to configure and schedule a **storage management** task to delete and manage old versions of data to free up data lake storage. This archival process is driven by inexpensive storage and lends itself to the data lake concept.
- TimeXtender's ODX can also automatically **move** old versions in the data lake from hot to cool storage to **save costs**; no manual work needed!
- TimeXtender makes it super easy to create & maintain connections to hundreds of types of different data sources. Using the "Data Source Wizard" users can choose from **260+ data source connectors**, enter their credentials and immediately begin the synchronization. Compared to 90 different connectors in Azure Data Factory, the TimeXtender way makes it far easier to connect and maintain your connections.
- TimeXtender lets you create an always up-to-date, **full documentation** of your complete data lake with a few clicks.
- TimeXtender also provides you with visual **data lineage** and impact analysis so that you can track where the data is being used in the entire data estate.

**A hand-built data lake will not have any of this functionality.**



# CONCLUSION

A data lake can now be easily and quickly built using automated data integration tools like TimeXtender. TimeXtender also provides a number of other benefits, such as the ability to automatically connect to any data source, create full documentation of the data lake, and track data lineage.

This means that data engineers can free themselves from these tedious, manual tasks, and focus on more important tasks, such as data modeling and data analysis.



If you're looking for a data integration tool that can help you build a future-proof data lake 10x faster, TimeXtender is the solution for you. To learn more, please visit [www.TimeXtender.com](http://www.TimeXtender.com)

