



Getting The Most From Your Azure Migration with Data Mastering



Microsoft Azure

Migration Guide

Getting The Most From Your Azure Migration with Data Mastering

The benefits of cloud technology are a true business game-changer, which is why many organizations are moving to Azure. But before you shift to the cloud, you'll need to reconcile how to best migrate your data in order to take full advantage of the cloud's benefits.

The good news about an impending cloud migration is that you have a chance to start fresh. Now is an excellent time to address your legacy data problems and transform your business' data into an asset and make it readily available to the entirety of the business for downstream efforts like data science, analytics, and business forecasting.

Here's the bad news: if you're counting on a simple, straight-forward lift and shift migration strategy, you're setting yourself and your organization up for failure. The reason? With a lift and shift strategy all the issues that plagued it when hosted on-premise (think siloed data, duplicate records, incomplete records) will follow you to the cloud. Your data will be just as unusable as it was before on-premise. This is why before any migration, it is critical to think about how you will clean, curate, and master your data.

A cloud migration strategy that is optimized for ROI not only moves, but *improves* data.

Discover Cloud-Native
Data Mastering

[START MASTERING](#)

What Your Move to Azure Shares with Moving to a New Home

A good way to think about your upcoming Azure migration is like moving into a new home. At the moment all your stuff (ie., data) is at your current home in different states of organization. When prepping to move, would you pack up the messes in your old place, and simply ship the messes to your new place? Probably not. You're more likely to neatly organize and pack the things you plan to bring so that when it arrives at your new place you know what you have, it's easy to unpack, and it's organized giving you a fresh start.

The same logic applies to a cloud migration; but with one massive advantage. Because cloud compute power and storage is far more economical than on-premise, by migrating to the cloud not only can you store data more cheaply, but you can activate computationally intensive machine learning algorithms to do the majority of organizing, enriching and mastering at the same time. In essence, outsourcing the hard work of cleaning your bad data to a machine while it's in transit: a major jump in speed and efficiency. **By taking advantage of machine learning in the cloud, it's possible to manage 10 times as much data, with one-tenth the people and in one-tenth the time.**

Tamr's cloud-native data mastering solution uses machine learning to do the heavy lifting of curating and enriching data, so your organization can use the data in the cloud to drive radically better business decision making and real business outcomes from mastered data - saving money, driving growth and reducing risk.

With the need for clean, curated mastered data prior to a cloud migration established, the questions are how and when do you do this? Modern data mastering possess these critical features:

1. Machine learning to master data at hyper-scale

Traditionally, organizing and mastering data has been done with a rules-based approach (if / then). Conventional rules-based systems can be effective on a small scale, relying on human-built rules logic to generate master records. However, rules quickly fall apart when tasked with connecting and reconciling large amounts of highly variable data at scale. Machine learning, on the other hand, becomes more effective at matching records across datasets as more data is added. In fact, huge amounts of data (1M+ records across dozens of systems) provides more signals for the algorithms to identify patterns, matches, and relationships, accelerating years of human effort down to days.

2. Open and interoperable architecture to break down existing data silos

Look for a solution with an open and interoperable architecture that allows businesses to pursue “best-in-breed” solutions for all their data needs. Today’s premier data organizations take a DataOps approach to their technology stacks, which means using the best tool for each specific need, instead of what’s easiest or readily available. Look for solutions that play well with others and are complementary through RESTful APIs and robust integration capabilities.

3. Cloud-native technologies that scale effectively

Machine learning is essential to improving data quality. As stated before, manual, rules-based approaches don’t scale and are slow to provide value. However, running large machine learning projects on-prem is incredibly costly and computationally taxing. This is where the cloud can make all the difference. The cloud provides the scale and compute that makes using machine learning efficient and cost-effective.

Additionally, cloud-native solutions are ideal for leveraging the flexibility and scalability of Azure. Cloud-native capabilities (technologies that leverage built-in elastic and ephemeral cloud and compute benefits of cloud technology) allow for a highly secure and scalable infrastructure that is able to add additional storage and compute power without adding to physical and hosting costs. With this built-in advantage, cloud-native solutions allow organizations to reduce the total cost of ownership and enable data organizations to take advantage of ongoing product enhancements and tooling without needing to allocate additional resources to hardware, and system or software upgrades.

Moving Day: Choosing When and Where to Master Data

There's also the choice of when in the data migration flow to master your data. An initial thought may be to master everything prior to moving data into the cloud—really leaning into the idea of starting clean. However, there are plenty of advantages to mastering data once it is staged in the cloud. Consider the positives and negatives of each approach in the table below.

Master Data On-Premise



Data is mastered before entering Azure making it valuable to the entire business



Data may still be siloed and unavailable to the mastering effort



Costly to establish a short-term environment to run large-scale machine learning algorithms on large data sets. This effort can be focused on establishing capabilities in the new cloud environment



Still need to move data to the cloud



Need to work with additional technologies to connect to on-prem sources and move data to the cloud on either side of the mastering process

Master Data in the Azure Cloud Data Lake



Data is mastered before entering Azure making it valuable to the entire business



Improved data access allowing Tamr to be applied to the entire corpus of data. This can help identify data sources that are redundant and should proceed no further in the migration workflow



Once provisioned, the cloud-native Tamr instance can continually master data as new sources are added to the data lake both now and in the future



Data is already in the cloud



Data is already in the cloud. Can take advantage of established data migration patterns and read data directly from the cloud data lake

Case study: How a Top 3 Global Shipping Company Turns Customer Data into Upsell Opportunities with Tamr

Our customer, a leading global shipping company based in Europe, wanted to create 360-degree customer views and use these records to increase sales among its existing customer base. However, the rules-based master data management system they were using had several shortcomings, including:

- **Being slow to provide insight and value:** After five years, a team of 60 mastered only five of 130 entities.
- **Failure to scale:** A rules-based approach was unable to handle our customer's ever-growing volume of data.
- **Requiring additional resources:** Mastering additional entities would have required substantial headcount increases.

After a successful proof of concept that yielded more than 1 million mastered customer records in weeks by leveraging machine learning, the customer decided to replace its legacy MDM solution with Tamr. Tamr's cloud-native architecture also appealed to our customer since it would allow them to leverage the scalability and flexibility of Azure to master their data. Some of the business outcomes our customer has achieved include:

- **Driving upsell opportunities:** Customer upsell opportunities are valued at several million dollars.
- **Delivering faster time to value:** New data sources are onboarded in weeks compared to months. required substantial headcount increases.
- **Better use of analysts' time:** Drastically reduce the time analysts spent manually preparing data to let them focus on higher value projects.

Tamr and Azure: Offering a Modern, Agile Data Platform

Tamr is complementary to Azure and delivers mastered data to components like Synapse and Purview to support downstream analytics using services such as PowerBI and Azure ML Studio. Used together, Tamr and Azure help enterprises form a modern, agile data platform.

Azure Data Catalog: Azure Data Catalog is a knowledge source for Tamr and provides data that can be integrated, what does it contain and who uses it. Tamr also uses Azure Data Catalog as a registry that's updated with clean, mastered data for each entity. When users need a trusted source of data, they can search Azure Data Catalog and for items tagged as mastered by Tamr.

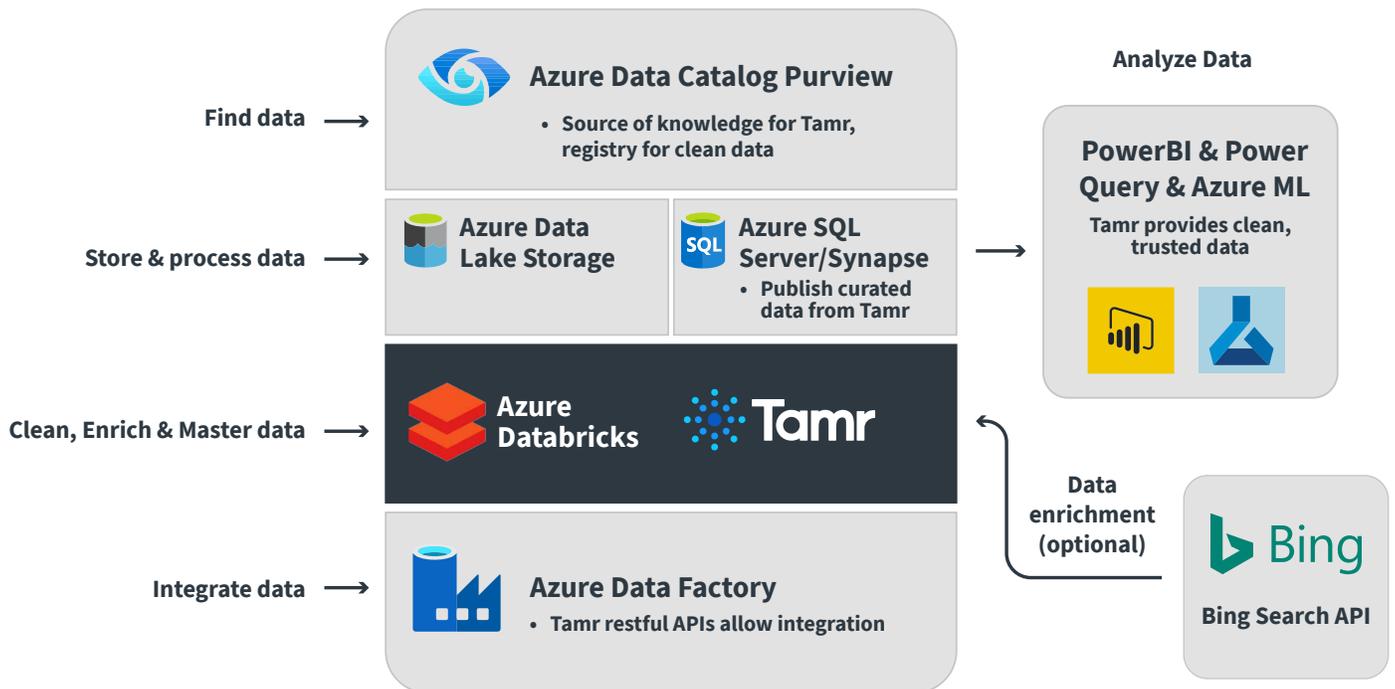
Azure Blob or Data Lake Storage: Data mastered with Tamr can be published to either Azure Blob or Data Lake Storage, providing users with access to high-quality, curated data.

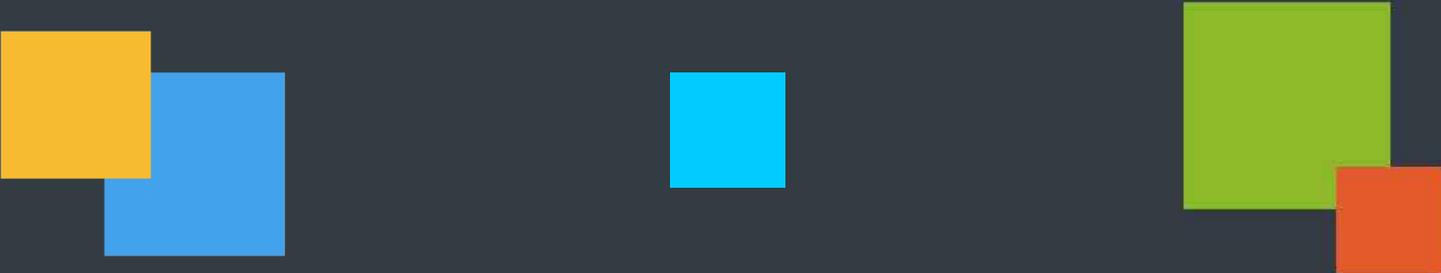
Azure Databricks: Tamr lets organizations use the cloud-native capabilities of Azure to scale their use of services like Azure Databricks, which Tamr uses for compute. Leveraging the elastic and ephemeral capabilities of Azure to increase and decrease compute as needed drives cost-efficient data mastering.

Azure Data Factory: Integration via Tamr's RESTful APIs allow data quality workflows to be easily invoked as part of a data pipeline. This reduces the barrier to applying data mastering processes.

PowerBI / Azure ML Studio: Analytics and machine learning tools like PowerBI and Azure ML Studio need clean, trusted data, which Tamr provides in a reliable and repeatable manner.

Bing Search API: The enrichment services offered through Bing Search API can be an additional source of data to process using Tamr. Enriching data often involves overcoming challenges around data variety, which is a core problem that Tamr solves.





Closing Thought

One of the most valuable assets unlocked by moving to the cloud is the speed at which data can be utilized to solve end-business problems. But doing this relies upon data being mastered.

Migrations are the perfect catalyst for a conversation around improving data quality. To get the most from a migration, lead with data.

By coming to the cloud with a solid understanding of your critical data, like how many customers you have, who are your leads, how many suppliers, or what parts you buy, you now have a known baseline to plug into existing applications today and new ones tomorrow; setting yourself up for success.

Using Tamr's machine-learning, interoperable, modular approach to mastering data, your migration can amplify the productivity and possibilities of your data in its new home. With Tamr, Azure customers can improve their migrations and be in a position to accelerate critical analytical insights by reconciling internal and external data at scale.

Learn more about how Tamr and Azure help organizations overcome their toughest data challenges.

[TAMR + MICROSOFT AZURE](#)