Microsoft

Empower IT and Data Professionals to Achieve More with All Their Data

A guide to migrating existing processes to Azure Synapse Analytics

Introduction

Many companies today have long had analytical data warehouses in place in their data centers to support decision making in different parts of their business. Sales, marketing, and finance departments in particular are heavy users of such systems producing standard reports and dashboards. They also employ business analysts to perform ad hoc query and analysis on data in data marts designed for multi-dimensional analysis using self-service business intelligence (BI) tools.

However, while data warehouses support some key aspects of decision making, the shift to digital technologies over the last several years has seen changes happening around the data warehouse that are having a major impact on traditional analytical systems. This includes source transaction processing systems–such as CRM, HR, and ERP systems–moving to the cloud. Often this is done to consolidate similar systems, to scale and to enable mobile self-service access to transaction processing systems by customers (e.g. mobile banking), partners, suppliers, and employees. This has resulted in data warehouses now needing to capture data from cloud SaaS applications. In addition, new data has emerged that businesses are now capturing such as on-line clickstream, social network data, internet of things (IoT) sensor data, open government data, weather data, image, audio and video data. However, very little of this type of data has made it into data warehouses. In fact, it is often processed and analyzed separately at scale in the cloud while data scientists use it to build machine learning predictive and prescriptive analytics.

In addition, with so many new data and analytics technologies now appearing on the cloud first, the attraction of moving analytical workloads to the cloud to quickly take advantage of these technologies is becoming more and more compelling. Given this backdrop, **it is not surprising that many companies are now considering moving their data warehouses to the cloud as part of a data warehouse modernization exercise.**

Today, many organizations are migrating their legacy data warehouse solutions to Azure Synapse Analytics to gain the benefits of and end-to-end analytics platform that provides high availability, security, speed, scalability, cost savings, and industry-leading performance for enterprise data warehousing workloads.

As technologies evolve, the benefits of having a cloud-based data warehouse solution far outweigh its on-premises counterparts. Azure Synapse not only provides industry-leading performance for running enterprise data warehousing workloads in the cloud, it is an end-to-end analytics platform that brings data ingestion, data warehousing, and big data analytics into a single service. With its scalability and separate compute and storage architecture, Azure Synapse can be scaled instantly in ways not possible with legacy systems such as Teradata, Netezza, or Exadata.

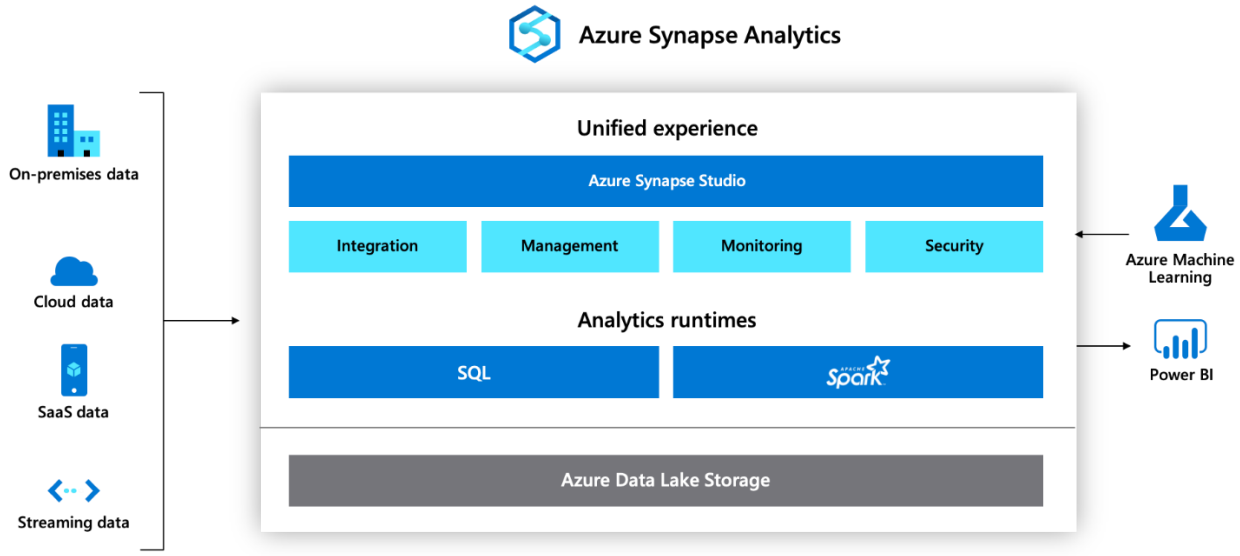
There are several business benefits as to why you should consider migrating to Azure Synapse to help you lower total cost of ownership, improve price/performance, and leverage a rich ecosystem of additional data and analytical technologies that can help you modernize your data warehouse and also shorten time to value.

Business benefits include:

- Lower implementation and maintenance costs – pay only for what you use
- No infrastructure to manage so you can focus on competitive insights
- Pay for data and analytics tools only when needed, pausing consumption when not in use

- Reduction in analytics project development time and increased innovation
- Ability to independently scale storage from compute
- Low cost storage for staging and production data
- Avoid expensive upgrades as data volumes grow and capacity is used up by ELT processing
- Lower storage costs as your data volumes grow
- Higher levels of security and business continuity
- Accelerated time to insight with an analytics platform that natively integrates Apache Spark & SQL engines
- Future proof architecture that easily incorporates newest innovations

Beyond the business benefits of running your enterprise data warehouse in the cloud with Azure Synapse, all data professionals can now use a unified experience for data prep, data management, data warehousing, big data, and AI tasks. With two ways to analyze data through provisioned workloads or through the severless consumption model that offers pay-per-query functionality, customers can choose the most cost-effective option for each use case. Additionally, when it comes to data, security and privacy are of utmost importance and underpin the insights discovered using analytics. Advanced security and privacy are built into the fabric of Azure Synapse, such as always-on data encryption. For fine-grained access control, businesses can help ensure data stays safe and private using column-level security and native row-level security, as well as dynamic data masking to automatically protect sensitive data in real-time.



This guide provides the high-level methodology for planning, preparing, and executing a successful migration of an existing legacy data warehouse system to Azure Synapse Analytics. This guide is not intended to be a comprehensive step-by-step manual for migration, but rather a practical overview to help with your migration planning and project scoping.

The appendix at the end of this guide identifies some of the common migration issues and possible resolutions.

The intended audiences of this guide are data warehouse architects, solutions architects, CTOs, and project managers who need a clearly defined approach for migrating an existing on-premises data warehouse to Azure Synapse Analytics.

Why migrate

Data Warehouse Migration process

3-step migration process (prep, migrate, post)

2 types of migration (Lift & shift, Re-design)

Reduce complexity before migrating

Migrating existing schema

Migrating historical data

Migrating existing ETL

Migrating BI/queries

Security considerations & Tooling

Conclusion

Why should you migrate your legacy data warehouse to Azure Synapse Analytics?

By migrating to Azure Synapse Analytics, companies with legacy data warehouse systems can take advantage of the latest innovations in cloud technologies and delegate tasks such as infrastructure maintenance and platform upgrade to Azure.

Customers who have migrated to Azure Synapse are already reaping many of its benefits, including:

Performance

Azure Synapse Analytics offers the best-of-breed relational database performance by using techniques such as Massively Parallel Processing (MPP) and automatic in-memory caching. The results of this can be seen in independent benchmarks such as the one by GigaOm (reference: https://gigaom.com/report/data-warehouse-cloud-benchmark/). This report compares Azure Synapse to other popular cloud data warehouse offerings.

Speed

Data warehousing is process intensive. It involves data ingestion, transforming data, cleansing data, aggregating data, integrating data, and producing data visualization and reports. The many processes involved in moving data from original sources to a data warehouse are complex and interdependent. A single bottleneck can slow the entire pipeline and an unexpected spike in data volume amplifies the need for speed. When timeliness of data matters, Azure Synapse Analytics meets the demand for fast processing.

Improved security and compliance

Azure is a globally available, highly scalable, secure cloud platform. Azure Synapse Analytics, which resides inside the Azure ecosystem, inherits all of the aforementioned goodness.

Elasticity and cost efficiencies

In a data warehouse, the demands for workload processing can fluctuate. At times, these fluctuations can vary drastically between peaks and valleys. For example, sudden spikes in sales data volume can occur during the holiday seasons. Cloud elasticity allows Azure Synapse to quickly increase and decrease its capacity according to demands with no impact to infrastructure availability, stability, performance, and security. Best of all, you only pay for your actual usage.

Managed infrastructure

Eliminating the overhead of data center management and operations for the data warehouse allows companies to reallocate valuable resources to where value is produced and focus on using the data warehouse to deliver the best information and insight. This lowers overall total cost of ownership and allows for better cost control over your operating expenses.

Scalability

The volume of data in a data warehouse typically grows as time passes and as history is collected. Azure Synapse Analytics can scale to match this growth by incrementally adding resources as data and workloads increase.

Cost savings

Running an on-premises legacy data center is expensive (costs of servers and hardware, networking, physical room space, electricity, cooling, and staffing). These expenses can be substantially minimized with Azure Synapse Analytics.

Azure Synapse Analytics provides you with true pay-as-you-go cloud scalability without the need for complicated reconfiguration as your data or workloads grow.

Maximize skills

Bring all the existing skills across your business together to accomplish more with your data. With the deeply integrated Apache Spark and SQL engines in Azure Synapse, data professionals who prefer familiar SQL can seamlessly collaborate with those who prefer Spark—and vice versa.

For instance, those familiar or who prefer SQL can query Spark tables using the T-SQL language. And data engineers or data scientists who prefer languages such as Python, Scala, SparkSQL, or C# can transform data, train models, and create proofs of concept in the same service that houses data pipelines, data lakes, and data warehouses.

Data lake to data warehouse

Manage, secure, and analyze all types of data. Azure Synapse can query structured or semi-structured data with data warehousing resources, and also quickly execute a serverless query over unstructured data from your data lake. Enable your data professionals to build end-to-end analytics solutions without having to stitch a multitude of services together.

Take dissolving data silos to the next level with Azure Synapse Link—a cloud-native hybrid transactional analytical processing (HTAP) implementation now available in public preview. This technology removes the barriers between Azure database services and Azure Synapse—enabling customers to get insights from their live transactional data stored in their operational databases with a single click, without managing data movement or placing a burden on their operational systems.

Data Warehouse Migration Process

A successful data migration project starts with a well-designed plan. An effective plan accounts for the many components that need to be considered, paying particular attention to architecture and data preparation. Below is the 3-step migration process plan.

The 3-step migration process

Preparation

- Define scope of what is to be migrated
- Build inventory of data and processes for migration
- Define data model changes (if any)
- Define source data extract mechanism
- Identify suitable Azure (and 3rd party) tools and services to be used
- Train staff early on the new platform
- Set up Azure target platform

Migration

- Start small and simple
- Automate wherever possible
- Leverage Azure built-in tools and features to reduce migration effort
- Migrate metadata for tables and views
- Migrate historical data to be maintained
- Migrate or refactor stored procedures and business processes
- Migrate or refactor ETL/ELT incremental load processes

Post-Migration

- Monitor and document all stages of the process
- Use experience gained to build a template for future migrations
- Re-engineer the data model if required (using new platform performance and scalability)
- Test applications and query tools
- Benchmark and optimize query performance

Two types of migration strategies

Start your migration planning by doing an assessment on your existing data warehouse to determine which migration strategy works best for your situation. There are two types of migration strategies to consider:

Lift and Shift Strategy

For the Lift and Shift strategy, the existing data model is migrated unchanged to the new Azure Synapse Analytics platform. This is to minimize the risk and the time required for migration by reducing the scope of changes to the minimum.

Lift and Shift is a good strategy for legacy data warehouse environments such as Netezza where:

- a single data mart is to be migrated, or
- the data is already in a well-designed star or snowflake schema, or
- there are immediate time and cost pressures to move to a modern cloud environment

Redesign Strategy

In scenarios where the legacy data warehouse has evolved over time, it might be essential to re-engineer to maintain the optimum performance levels or support new types of data. This could include a change of the underlying data model.

To minimize risk, it is recommended to migrate first using the Lift and Shift strategy and then gradually modernize the data warehouse data model on Azure Synapse Analytics using the Redesign strategy. A complete change in data model will increase risks because it will impact source to data warehouse ETL jobs and downstream data marts.

Reducing complexity of your existing legacy data warehouse before migrating

In the previous section, we presented the two migration strategies. As a best practice, during the initial assessment step, be cognizant of any possibility of simplifying your existing data warehouse and document them. The goal is to reduce the complexity of your existing legacy data warehouse system before the migration to make the migration process easier.

Here are some recommendations on how to reduce complexity of your existing legacy data warehouse:

Remove and archive unused tables before migrating

- Avoid migrating data that is no longer in use

Convert physical data marts to virtual data marts

- Minimize what you have to migrate
- Reduce total cost of ownership
- Improve agility

In the next section, we will take a closer look at why you should consider converting a physical data mart to a virtual data mart.

Converting Physical Data Marts to Virtual Data Marts

Prior to migrating your legacy data warehouse, consider converting your current physical data marts to virtual data marts. By using virtual data marts, you can eliminate physical data stores and ETL jobs for data marts without losing any functionality prior to migration. The goal of this is to reduce the number of data stores to migrate, reduce copies of data, reduce the total cost of ownership and improve agility. To achieve this, you will need to switch from physical to virtual data marts before migrating your data warehouse. You can consider this as a data warehouse modernization step prior to migration.

Disadvantages of Physical Data Marts

- Multiple copies of the same data
- Higher total cost of ownership
- Difficult to change as ETL jobs are impacted

Advantages of Virtual Data Marts

- Simplifies data warehouse architecture
- No need to store copies of data
- More agility
- Lower total cost of ownership
- Uses pushdown optimization to leverage the power of Azure Synapse Analytics
- Easy to change
- Easy to hide sensitive data

Migrating existing data warehouse schema to Azure Synapse Analytics

Next, plan how you will migrate the schema of your existing legacy data warehouse. Schema migration involves the migration of existing staging tables, legacy data warehouse and dependent data mart schema.

To help you understand the magnitude and scope of your schema migration, we recommend that you create an inventory of your existing legacy data warehouse and data mart.

Here is a checklist to help you collect the necessary information:

- ✓ Row counts
- ✓ Staging, Data Warehouse and Data Mart data size
  - Tables and indexes
- ✓ Data compression ratios
- ✓ Current hardware configuration
- ✓ Tables (including partitions)
  - Identify small dimension tables
- ✓ Data types
- ✓ Views
- ✓ Indexes
- ✓ Object dependencies
- ✓ Object usage
- ✓ Functions
  - Both out-of-the-box functions and UDFs
- ✓ Stored Procedures

- ✓ Scalability requirements
- ✓ Growth projections
- ✓ Workload requirements
  - concurrent users

With your inventory completed, you can now make decisions on scoping what schema you want to migrate. Essentially, there are four options for scoping your legacy data warehouse schema migration.

1. Migrate one data mart at a time

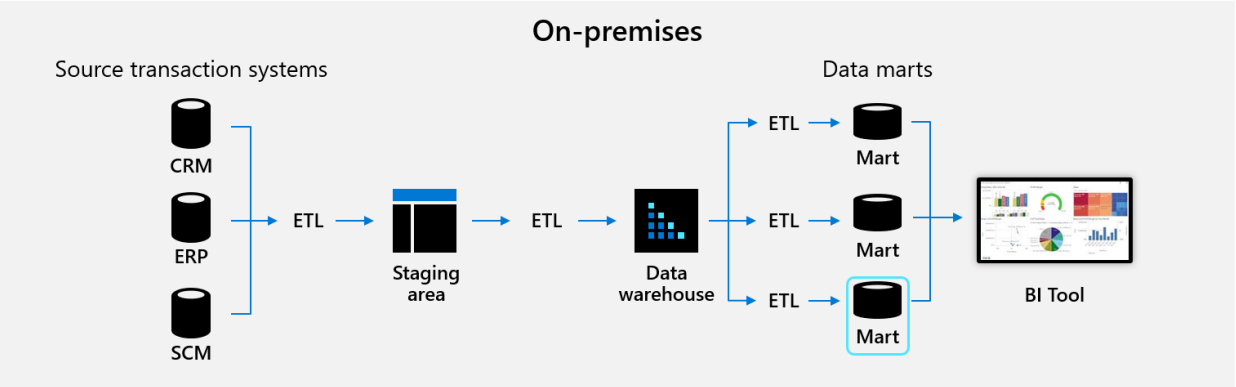


Figure 3 - Migrate one data mart at a time

2. Migrate all data marts at once, then the data warehouse

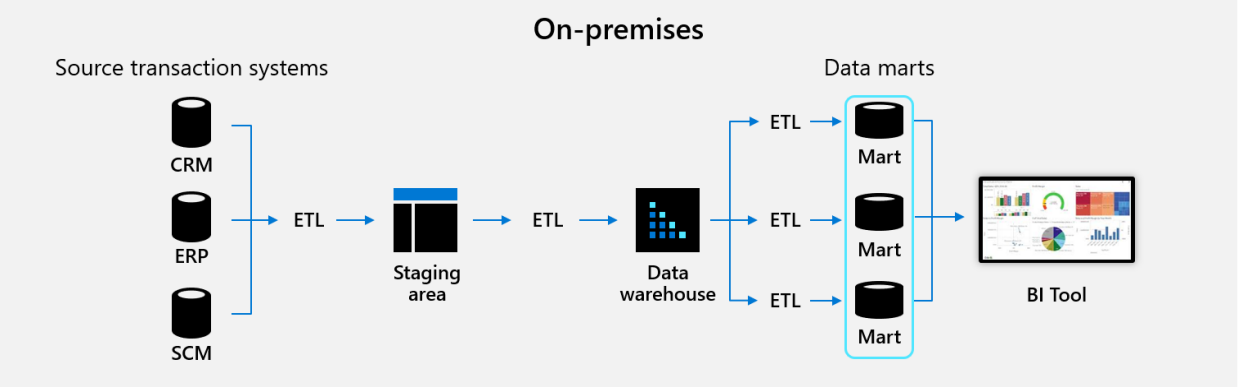


Figure 4 - Migrate all data marts at once, then the data warehouse

3. Migrate both data warehouse and staging area

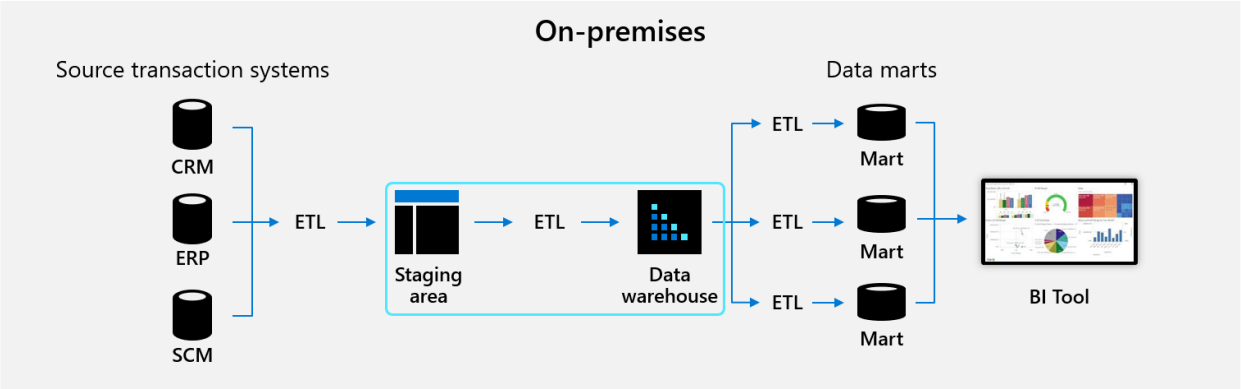


Figure 5 - Migrate both data warehouse and staging area

4. Migrate everything at once

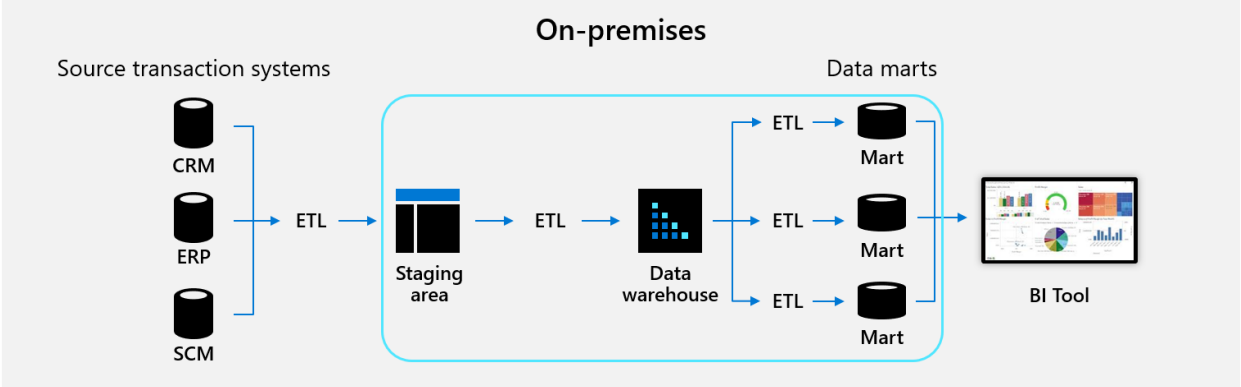


Figure 6 - Migrate everything at once

Keep in mind when choosing your option that the goal is to achieve a physical database design that will match or exceed your current legacy data warehouse system in performance and preferably at a lower cost.

To recap, here are some of the recommendations for the schema migration:

- Avoid migrating unnecessary objects or processes
- Consider using virtual data marts to reduce or eliminate the number of physical data marts
- Automate whenever possible
- Use Metadata from system catalog tables in the legacy data warehouse system to generate DDL for Azure Synapse Analytics
- Perform any required data model changes or data mapping optimizations on Azure Synapse Analytics

Migrating historical data and ETL processes from your legacy data warehouse to Azure Synapse Analytics

Once the schema migration scope has been determined, we are now ready to make decisions on how to migrate the historical data.

The steps for migrating historical data are as follows:

1. Create target tables on Azure Synapse Analytics
2. Migrate existing historical data
3. Migrate any functions and stored procedures required
4. Migrate incremental load (ETL/ELT) staging and processes for incoming data
5. Apply any performance tuning option required

The table below outlines the four data migration options and their pros and cons.

Data migration option	Pros	Cons
Migrate data mart data first followed by data warehouse data	• Migrating data from one data mart at time is an incremental low risk approach • Subsequent ETL migration is limited to only the data in the dependent data marts migrated	• Until your migration is complete, you will have some data that exist on-premises and on Azure • ETL processing from data warehouse to data marts would need to bridge the firewall and be changed to target Azure Synapse
Migrate data warehouse data first followed by data marts	• All data warehouse historical data migrated	• Leaving dependent data marts on-premises not ideal as ETLs would have to flow data back into the data center • No real opportunity for incremental data migration
Migrate data warehouse and data marts together	• All data migrated in one go	• Potentially higher risk • Means ETLs will most likely all have to be migrated together also
Convert physical marts to virtual marts and only migrate the data warehouse	• No data mart data stores to migrate • No ETLs from data warehouse to marts to migrate • Only the data warehouse data to migrate • Fewer copies of data • No loss in functionality • Lower total cost of ownership • More agility • Simpler overall data architecture • May be possible with views in Azure Synapse	• If nested views are not capable of supporting virtual data marts then 3rd party data virtualization software on Azure will likely be needed • All marts would need to be converted before data warehouse data is migrated • Virtual marts and data warehouse to virtual mart mappings will need to be ported to Data Virtualization server on Azure and redirected to Azure Synapse

Migrating existing ETL processes to Azure Synapse Analytics

There are a number of options available for migrating your existing ETL processes to Azure Synapse Analytics. The following table outlines some of the ETL migration options based on how the existing ETL jobs were built.

How are existing ETL jobs built?	Migration options	Why migrate and what to look out for
Custom 3GL code and scripts	• Plan to re-develop these using Azure Data Factory	• Code provides no metadata lineage • Hard to maintain if authors have gone • If staging tables are in the legacy data warehouse and SQL is used to transform data then resolve differences with T-SQL
Stored procedures that run in your legacy data warehouse DBMS	• Plan to re-develop these using Azure Data Factory	• Likely to be significant differences between legacy data warehouse and Azure Synapse • No metadata lineage
Graphical ETL tool (e.g. Informatica, Talend, etc.)	• Continue using your existing ETL tool and switch the target to Azure Synapse • Possibly move to an Azure version of your existing ETL tool and port the metadata to run ELT jobs on Azure making sure you enable access to on-premises data sources • Control execution of ETL services using Azure Data Factory	• Avoids re-development • Minimizes risk and quicker to migrate
Data Warehouse Automation software	• Continue using your existing ETL tool switching the target and staging to Azure Synapse	• Avoids re-development • Minimizes risk and quicker to migrate

Re-developing scalable ETL processes using Azure Data Factory

Another option for handling your existing legacy ETL processes is by re-developing them using Azure Data Factory (ADF). ADF is an Azure data integration service for creating data-driven workflows (known as pipelines) for orchestrating and automating data movement and data transformation. You can use ADF to create and schedule pipelines to ingest data from different data stores. ADF can process and transform data by using compute services such as Spark, Azure Machine Learning, Azure HDInsight, Hadoop, and Azure Data Lake Analytics.

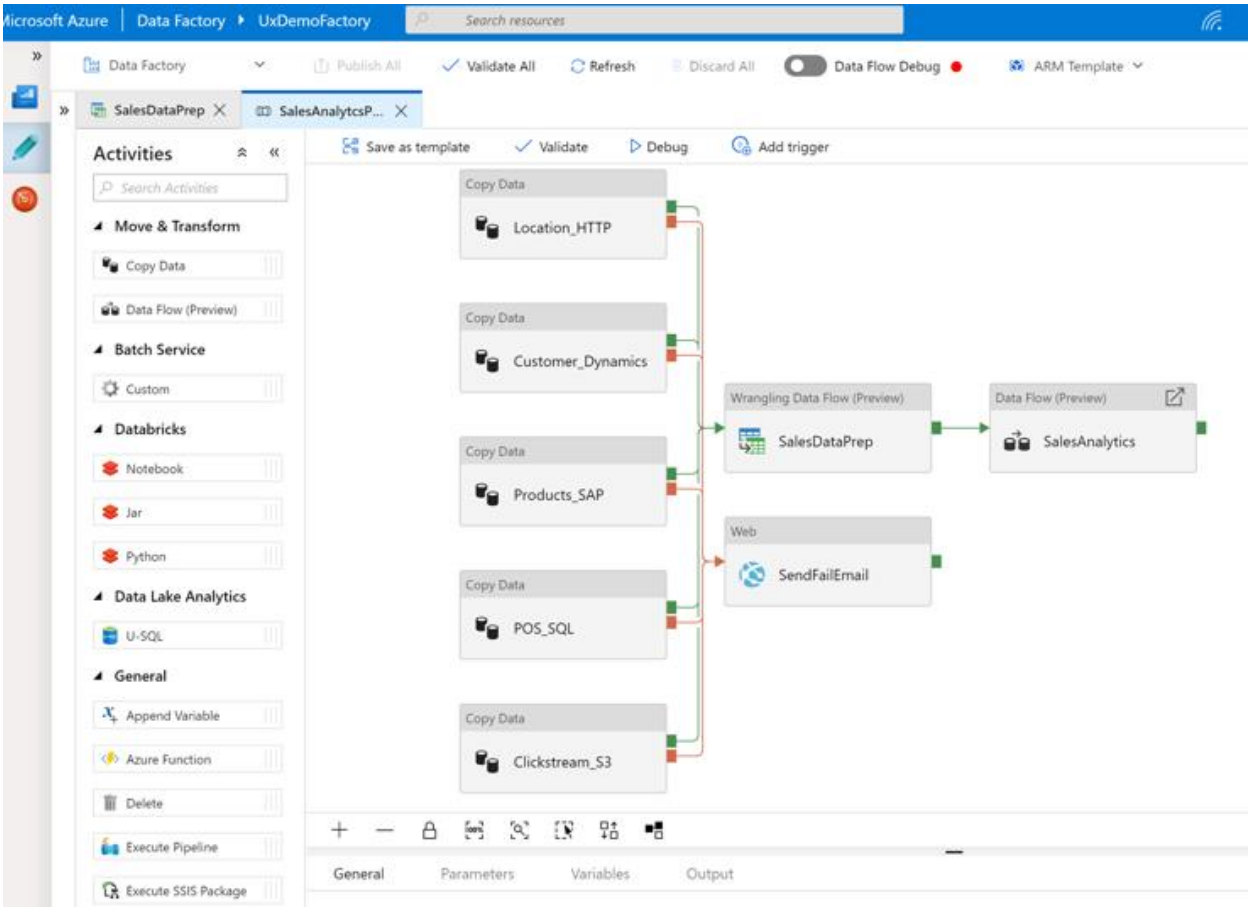


Figure 10 - Re-developing scalable ETL processes using Azure Data Factory

Recommendations for migrating queries, BI reports, dashboards, and other visualizations

Migrating queries, BI reports, dashboards, and other visualizations from your legacy data warehouse to Azure Synapse Analytics would be straight-forward if the legacy system uses standard SQL.

However, often times, this is not the case. In this situation, a different strategy must be taken:

- Identify the high priority reports to migrate first.
- Use usage statistics to identify which reports are never used.
- Avoid migrating anything that is no longer in use.
- Once you have produced the list of reports to migrate, their priorities, and which unused reports to bypass, confirm this list with the stakeholders.
- For reports that you are migrating, identify incompatibilities early to gauge the migration effort
- Some of the incompatibilities might be due to unsupported data types. See **Appendix A - Common migration issues and resolutions**.

- Consider data virtualization to protect BI tools and applications from structural changes to the data warehouse and/or data mart data model which might occur during the migration

Security considerations & Tooling

Protecting and securing your data assets is paramount in any data warehouse system. When planning a data warehouse migration project, security, user access management, backup, and restore must also be taken into consideration. For instance, data encryption may be mandatory for industry and government regulations such as HIPAA, PCI and FedRAMP as well as in non-regulated industries.

Azure includes many features and functions as standard that would traditionally have to be custom-built in legacy data warehouse products. Azure Synapse supports data encryption at rest and in motion as standard:

Data at rest

- Transparent Data Encryption (TDE) can be enabled to dynamically encrypt and decrypt Azure Synapse data, logs, and associated backups.
- Azure Data Storage can also automatically encrypt non-database data.

Data in motion

- All connections to Azure Synapse Analytics are encrypted by default, using industry standard protocols such as TLS and SSH

In addition, Dynamic Data Masking (DDM) can be used to obfuscate data for given classes of users based on data masking rules.

As a best practice, if your legacy data warehouse contains complex hierarchy of permissions, users and roles, consider using automation techniques in your migration process. You can use existing metadata from your legacy system to generate the necessary SQL to migrate users, groups, and privileges on Azure Synapse Analytics.

Tools to help migrate to Azure Synapse Analytics

Now that we have covered the planning, preparation and an overview of the migration process, let's have a look at the tools that you can use for migrating your legacy data warehouse to Azure Synapse Analytics. The tools that we will discuss are:

- Azure Data Factory (ADF)
- Azure Data Warehouse Migration Utility
- Microsoft Services for Physical Data Transfer
- Microsoft Services for Data Ingestion

Azure Data Factory (ADF)

- ADF is a fully managed, pay-as-you-use, hybrid data integration service for cloud scale ETL processing
- Processes and analyzes data in memory and in parallel to scale and maximize throughput

- Create data warehouse migration pipelines that orchestrate and automate data movement, data transformation, and data loading into Azure Synapse Analytics
- Can also be used to modernize your data warehouse by ingesting data into Azure Data Lake, processing and analyzing data at scale and loading into a data warehouse
- Supports role-based user interfaces for mapping data flows for IT professionals and self-service data wrangling for business users
- Can connect to multiple data stores spanning data center, clouds and SaaS applications
- Over 90+ natively built and maintenance-free connectors available (reference: https://azure.microsoft.com/en-ca/services/data-factory/)
- Mix and match wrangling and mapping data flows in the same pipeline to prepare data at scale
- ADF orchestration can control data warehouse migration to Azure Synapse Analytics
- Can execute SSIS ETL packages from Azure Data Factory

Azure Data Warehouse Migration Utility

- Migrate data from an on-premises SQL-Server-based data warehouse to Azure Synapse.
- Uses a wizard-like approach to perform a lift-and-shift migration of schema and data from an on-premises, SQL-Server-based data warehouse.
- You can select the on-premises database containing the table(s) that you want to export to Azure Synapse. Then you select the tables that you want to migrate and migrate the schema.
- Automatically generates T-SQL code needed to create an equivalent empty database and tables on Azure Synapse. Once you provide connection details to Azure Synapse you can run the generated T-SQL to migrate the schema.
- Following schema creation, you can use the utility to migrate the data. This exports the data from your on-premises SQL Server based data warehouse and generates BCP (Bulk Copy Program) commands to load that data into Azure Synapse.

Microsoft Services for Physical Data Transfer

Azure ExpressRoute

- Private connection between Azure and customer data centers
- Data does not go over the internet

AzCopy

- Copy data to Azure via the Internet

Azure Databox

- Large volumes (10's TB - 100's TB)

Microsoft Services for Data Ingestion

PolyBase (recommended method)

- Provides the fastest and most scalable bulk data loading into Azure Synapse Analytics
- Uses parallel loading to give the fastest throughput
- Can read from flat files in Azure Blob Storage or from external data sources via connectors
- Tightly integrated with Azure Data Factory

- CREATE TABLE AS or INSERT … SELECT
- Define staging table as type HEAP for fast load
- Support rows up to 1 MB in length

BCP (Bulk Copy Program)

- Supports rows larger than 1 MB in length
- Originally developed for earlier versions of Microsoft SQL Server
- Can be used to import and export data from any SQL Server environment including Azure Synapse Analytics
- (reference: https://docs.microsoft.com/en-us/sql/tools/bcp-utility)

SqlBulkCopy API

- This is an API equivalent of the BCP functionality
- Allows the implementation of load processes programmatically
- (reference: https://docs.microsoft.com/en-us/dotnet/api/system.data.sqlclient.sqlbulkcopy)

INSERT and INSERT … SELECT

- Azure Synapse Analytics supports standard SQL
- Load individual rows or results of SELECT statements into data warehouse tables
- INSERT … SELECT can be used within PolyBase to bulk insert data from extracted data via external data sources into data warehouse tables

Conclusion

Successful data migration projects start with a well-designed plan. An effective plan accounts for the many components that need to be considered, paying particular attention to architecture and data preparation.

Azure Synapse Analytics is a limitless analytics service with unmatched time-to-insight that accelerates the delivery of BI, AI, and intelligent applications for enterprises. You will gain a lot of benefits by migrating your legacy data warehouse to Azure Synapse Analytics, including performance, speed, improved security and compliance, elasticity, managed infrastructure, scalability and cost savings.

This guide provided the high-level methodology needed to prepare for and execute the migration of an existing Netezza system to Azure Synapse Analytics.

We have covered the 3-step migration process, the migration strategies, learned how to reduce complexity of your existing legacy data warehouse before migrating, and how to migrate existing schema, historical data, ETL processes and visualizations to Azure Synapse Analytics. We have also discussed the security considerations and tools which will help you succeed in your migration to Azure Synapse Analytics.

After you have migrated to Azure Synapse, you can explore additional Microsoft technologies in the rich Azure analytical ecosystem to modernize your data warehouse.

Here are some ideas to ponder:

- Offload your staging areas and ELT processing to Azure Data Lake and Azure Data Factory

- Build trusted data products once in common data model format and consume everywhere – not just in your data warehouse
- Enable collaborative development of data preparation pipelines by business and IT using ADF mapping and wrangling data flows
- Build analytical pipelines in ADF to analyze data in batch and real-time
- Build and deploy machine learning models to add additional insights to what you already know
- Integrate your data warehouse with live streaming data
- Simplify access to data and insights in multiple Azure analytical data stores by creating a logical data warehouse using PolyBase

All the best with your migration journey!

To learn more:

- Sign up for an Azure free account
- Connect with an Azure sales specialist on pricing, analytics best practices, setting up a proof of concept, and more.
- Learn why customers are choosing Azure for their analytics.

Get started today

Get started with an Azure free trial account:
https://azure.microsoft.com/en-us/free/synapse-analytics/

Download the free Getting Started Toolkit:
https://azure.microsoft.com/en-us/resources/azure-synapse-analytics-toolkit/

Learn more with the Azure Synapse documentation:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/

Appendix A - Common migration issues and resolutions

During the migration process, you might encounter certain issues that you need to overcome. In this section, we will highlight some of the common issues and provide you with resolutions that you can implement.

Issue #1: Unsupported data types and workarounds

The following table shows the data types from legacy data warehouse systems which are unsupported as well as the suitable workaround for Azure Synapse Analytics.

Unsupported data type	Workaround for Azure Synapse Analytics
geometry	varbinary
geography	varbinary
hierarchyid	nvarchar(4000)
image	varbinary
text	varchar
ntext	nvarchar
sql_variant	Split column into several strongly typed columns.
table	Convert to temporary tables.
timestamp	Rework code to use datetime2 and the CURRENT_TIMESTAMP function.
xml	varchar
user-defined type	Convert back to the native data type when possible

Issue #2: Integrity constraint differences

Pay close attention to the integrity constraint differences between your legacy data warehouse or data mart and Azure Synapse Analytics. In the diagram below, the left side represents the old legacy data warehouse system, and on the right side is the new Azure Synapse Analytics environment.

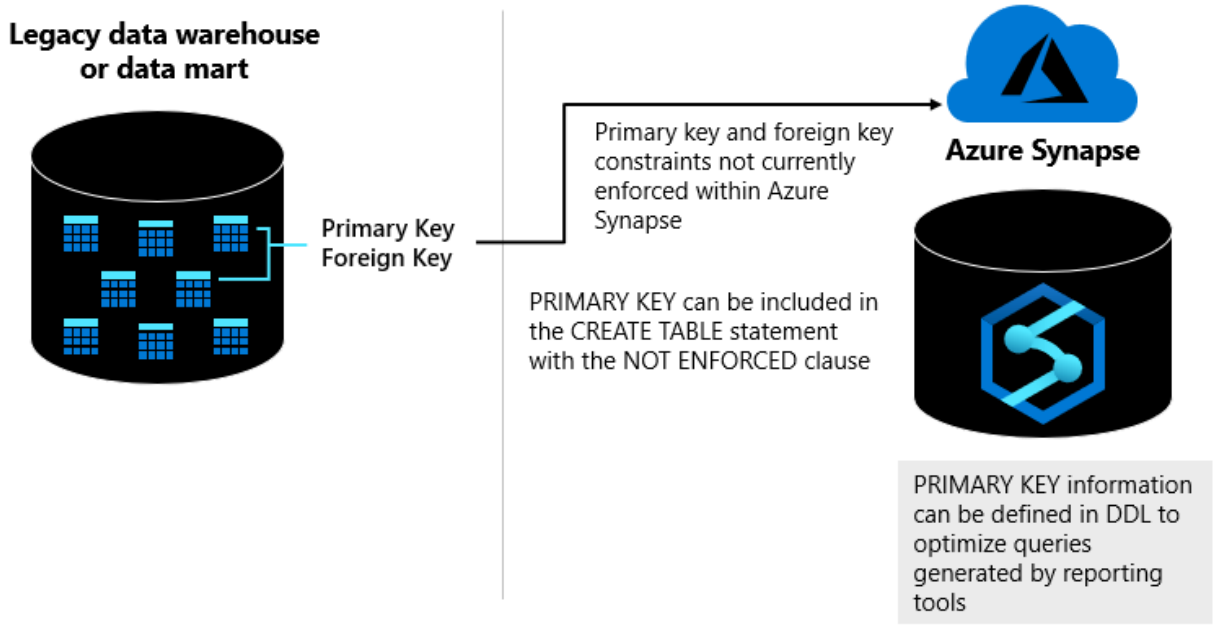


Figure 7 - Integrity constraint differences

In the subsequent sections, we will provide comprehensive coverage on how to resolve other common SQL incompatibles during the migration from legacy data warehouse to Azure Synapse Analytics.

Common SQL Incompatibilities and Resolutions

SQL Data Definition Language (DDL) differences and resolutions

Proprietary tables types

- On the legacy system, identify any use of proprietary table types
- Resolution: Migrate to standard tables within Azure Synapse Analytics
- For time series, index or partition on the date/time column
- Additional filtering will need to be added into the relevant temporal queries

Views

- On the legacy system, identify views from catalog tables and DDL scripts
- Views with proprietary SQL extensions or functions will have to be re-written
- Azure Synapse Analytics also supports materialized views and will automatically maintain and refresh these

Nulls

- NULL values can be handled differently in legacy SQL databases
  - For example, in Oracle, an empty string is equivalent to a NULL value
- Some DBMSs have proprietary SQL functions for handling NULLs
  - For example, NVL in Oracle
- Generate SQL queries to test for NULL values
- Test reports which include nullable columns

Extended SQL differences and workarounds

SQL Extension	Description	How to migrate
User Defined Functions	Can contain arbitrary code Can be coded in various languages (e.g. Lua, Java) Can be called within a SQL SELECT statement in the same way that built-in functions such as SUM() and AVG() are used	Use CREATE FUNCTION and re-code in T-SQL
Stored Procedures	Can contain one or more SQL statements as well as procedural logic around those SQL statements Implemented in a standard language (e.g. Lua) or in a proprietary language (e.g. Oracle PL/SQL)	Recode in T-SQL Some tools can help with migration E.g. Datometry, WhereScape
Triggers	Not supported by Azure Synapse	Equivalent functionality can be achieved by using other parts of the Azure ecosystem. For example, for streamed input data Azure Stream Analytics
In-database analytics	Not supported by Azure Synapse	Run advanced analytics such as machine learning models at scale is to use Azure Databricks Alternatively migrate to Azure SQL Database & use the PREDICT function
Geospatial Data Types	Not supported by Azure Synapse	Store geospatial data such as latitude/longitude and popular formats such as WKT (Well-Known Text) and WKB (Well-Known Binary) in VARCHAR or VARBINARY columns and accessed directly by geospatial client tools