[DOC] Transitioning from Export to Data Lake and BYOD

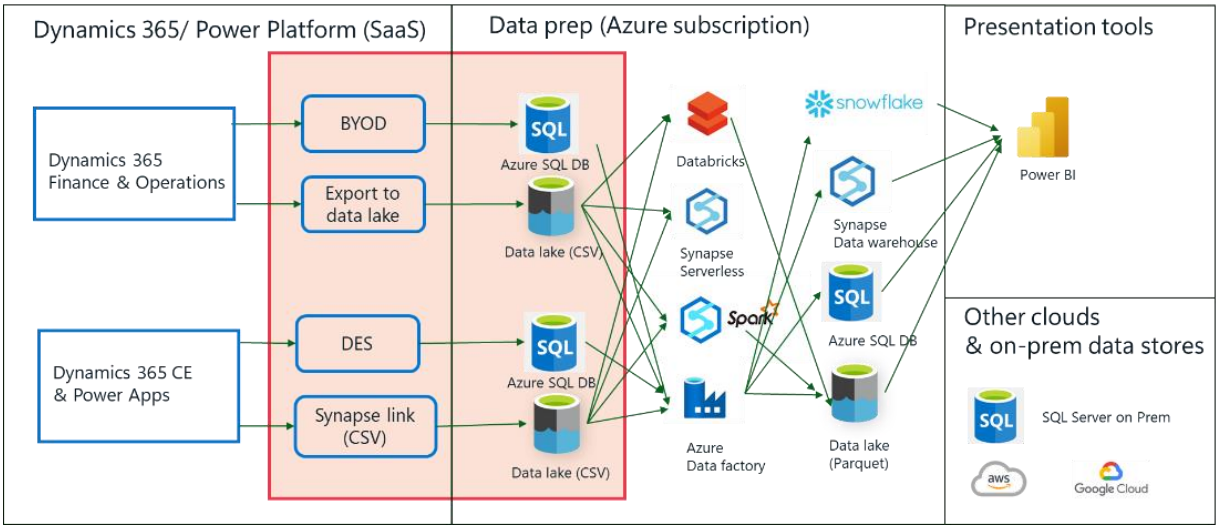
Data export service (DES), BYOD, Export to Data lake were features introduced in Dynamics 365 apps to export data for analytics and data integration scenarios. These services enabled IT admins and specialists to export data into external databases or data lakes and build data integration pipelines. While we have improved these services over the years with updates, as part of unification of Dynamics 365 with the power platform, we have rearchitected the same capabilities of these disparate services into simpler, unified experiences built into Power Apps maker portal. Transitioning to **Fabric Link** or upgrading to **Synapse Link**, the rearchitected services, provide you with an easy ramp to benefit from AI and Copilot investments in Dataverse and Microsoft Fabric.

If you are a customer using any of the previous generation services, this document provides guidance on upgrading to new experiences, benefiting from innovations as well as reducing end-to-end expenses and effort.

Based on preview customer surveys, we have also compiled a high level cost and benefits estimate to enable you to help with the transition. You will also see links to more information and videos as well as links to join forums and weekly office hour sessions to engage with product team, Microsoft specialists as well as fellow users as we continue to enhance these services with community participation.

Before

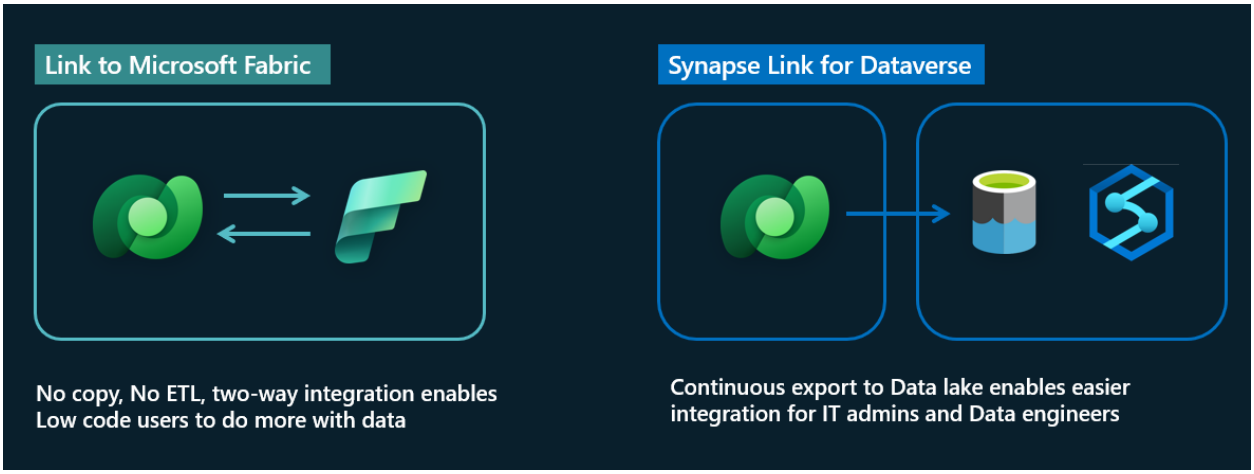
If you are a customer using legacy services BYOD, DES or Export to Data lake, you may have a data integration architecture similar to the one shown below. The highlighted box indicates the data pipelines your organization may have built to leverage the data exported from Dynamics 365 and Dataverse. You may use a selection of tools from Microsoft as well as others to copy and integrate Dynamics data with your own data. You may also transform and aggregate data by copying into multiple stores – shown under the Data-prep box. You may use Power BI or another tool to visualize data and create actionable insights. You may also have pipelines built to export data to an on-premise system and/or other clouds.



After transition

There are 2 data integration patterns enabled in Power Apps maker portal.

- **Synapse Link** enables continuous export of data similar to BYOD, Export to Data lake or DES services. This option is enabled for IT admins and data integration specialists.
- **Link to Fabric** feature provides a no-copy, no-ETL, fully managed software as a service (SaaS) integration.



These options are complementary and we provide a detailed comparison below.

If you are a Finance and Operations customer, currently using BYOD or Export to Data lake features, by upgrading to Synapse Link or Fabric Link, you can benefit from:

Simplified data integration architecture resulting in reduced operational costs

- Easy to configure and maintain via Power Apps maker portal - Built-in integration with Synapse and Microsoft Fabric

- Synapse Link and Fabric Link are fully managed services - requires minimal ongoing management overhead
- New services offer the same data shapes as previous services – your existing downstream integration pipelines can remain as is.
- Minimal impact to operational workloads, you don't need manage workloads and schedule data exports

Secure, end-to-end data integration pipelines

- With Fabric link, your data doesn't leave Dataverse governance boundary while authorized Fabric users can securely access data that resides within Dataverse
- Synapse Link service enables you to restrict access to your storage accounts with firewalls while enabling Dataverse to export data with Managed Service Identities (MSI) a security feature built into Microsoft Entra.

If you are a Dynamics 365 customer using Data Export service (DES) or classic Synapse Link with data exports in CSV format, you can benefit from efficient reporting enabled with the industry standard Delta/ Parquet data format.

- Built-in Delta/ parquet conversion option reduces the need to build your own pipelines for analytics and operational reporting
- Delta parquet format enables faster, more responsive queries & reports and scales to larger datasets of any size.
- Data in lake is compressed to 1/3 ~ 1/8 the original size - resulting in smaller files that reduce data query and carrying costs

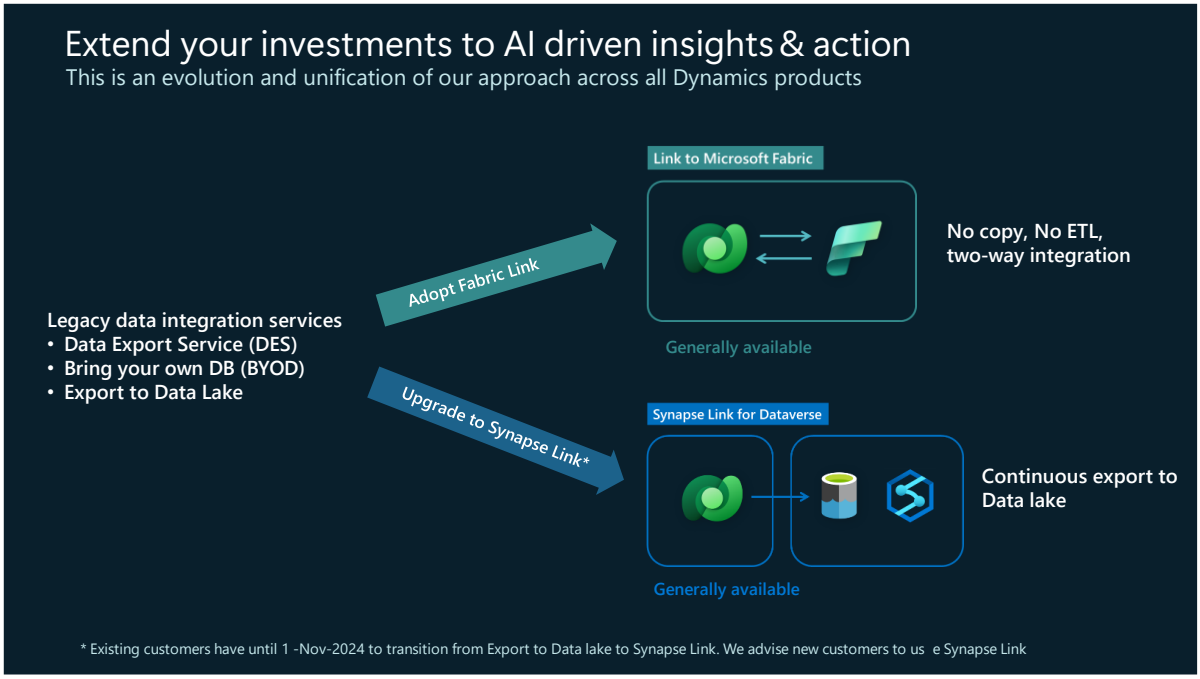
Comparison between Fabric Link and Synapse Link

A detailed comparison between Fabric Link and Synapse Link are shown below

Link to Fabric	**Azure Synapse Link**
No copy, no ETL direct integration with Microsoft Fabric.	Export data to your own storage account and integrate with Synapse, Microsoft Fabric, and other tools.
Data stays in Dataverse - users get secure access in Microsoft Fabric.	Data stays in your own storage. You manage access to users.
All tables chosen by default.	System administrators can choose required tables.
Consumes additional Dataverse storage.	Consumes your own storage as well as other compute and integration tools.

Which option should I use?

If your organization is already using Fabric or planning to transition, we recommend using Fabric Link feature. You can continue to use Synapse Link service if your immediate focus is to upgrade from your current services.



Simplification with Fabric Link

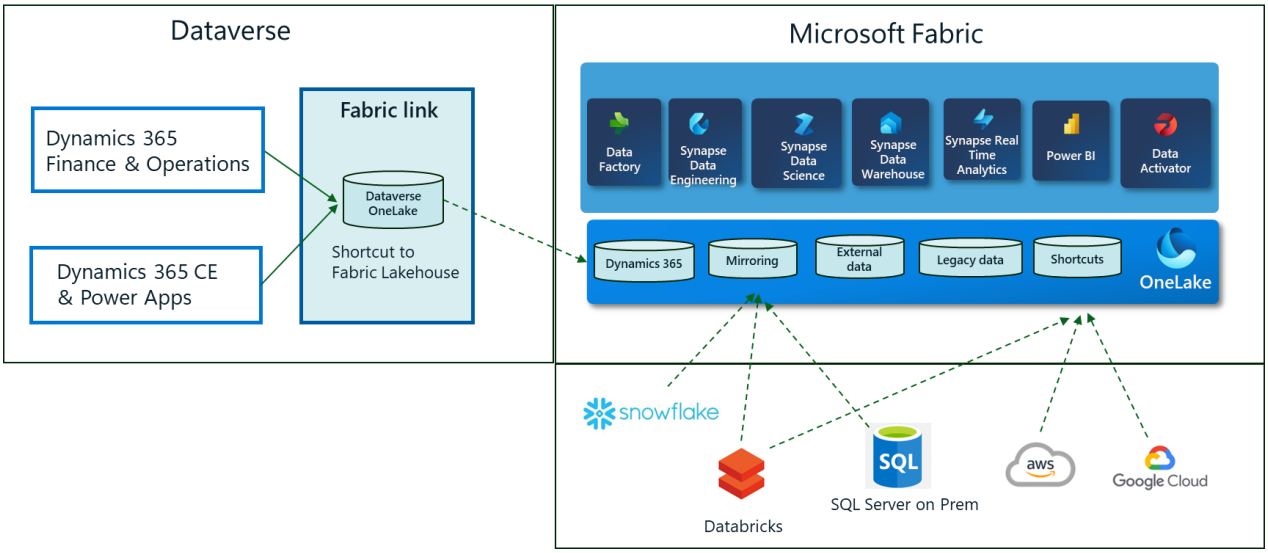
If you are already consuming data using Power BI, using Data warehouse, or using Data flows and Notebooks to transform data, Link to Fabric feature provides immediate value. You can simplify your data integration architecture by removing the need to have your own storage account or Synapse services for Dataverse data. Instead of paying for Azure resources like storage and compute, you will pay for the increase in Dataverse storage. Compute charges such as near real-time data updates and management overhead is also factored into Dataverse storage. Fabric Link option is like having a near real time read-only replica of your data optimized for insights.

You can query this replica using T-SQL, Spark/ python as well as all the workloads in Fabric. You can also access this data using any tool that can consume T-SQL as well as ADLS storage.

As a Dynamics or PowerApps customer, you get a Dataverse storage quota based on the number of licenses you purchased. Fabric Link feature uses this database quota. You can buy more storage add-ons if the data volume exceeds your quota.

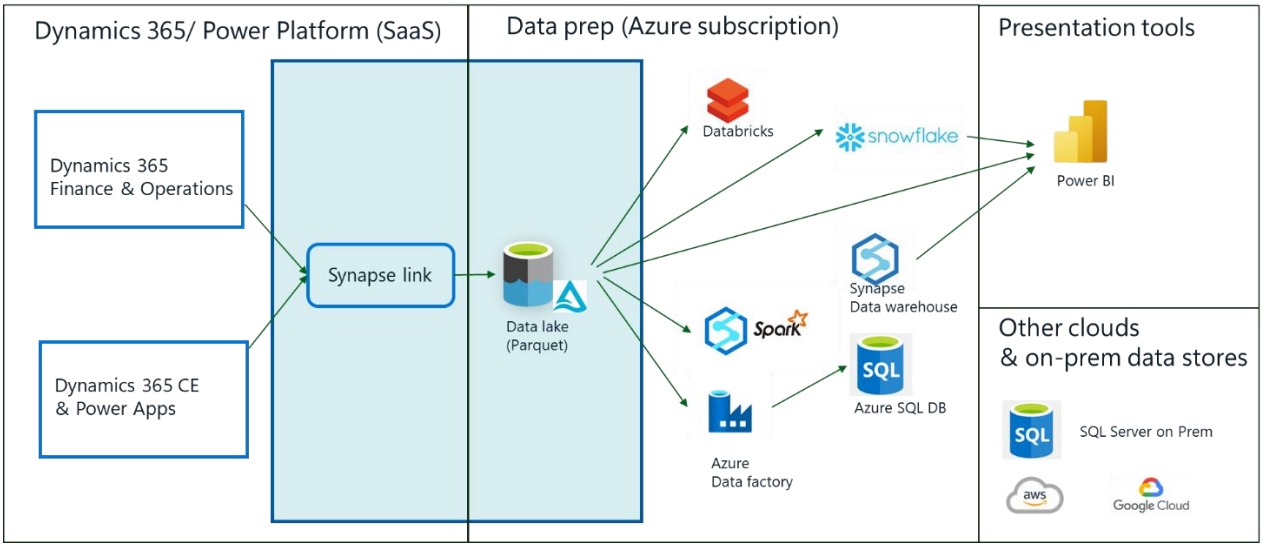
You can continue to retain Azure PaaS services like Databricks and SQL DBs in your own subscription. Recently announced Fabric features like data mirroring and shortcuts may help you further simplify your data integration.

Refer to examples on cost reductions achieved with the simplicity derived from Fabric integration feature below.



Upgrading to Synapse Link

By upgrading to Synapse Link and enabling delta parquet conversion, you can eliminate Dataverse data prep pipelines in your solution. Synapse Link service will export the same data shapes into your storage account in a more performant Delta/ parquet format. You can continue to use existing tools and Azure services like storage and Synapse query with minimal disruptions to your production environments.



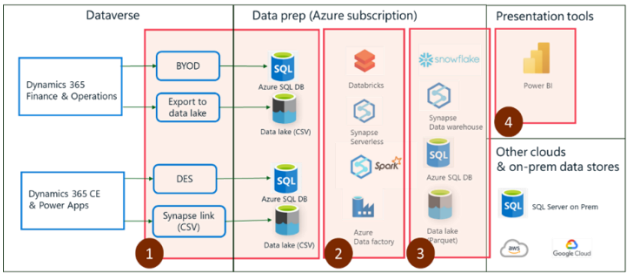
Understanding benefits – cost reductions

Simplicity achieved with Fabric Link and Synapse Link yields reductions in end-to-end costs. Consider the following examples that are based on actual customer experiences.

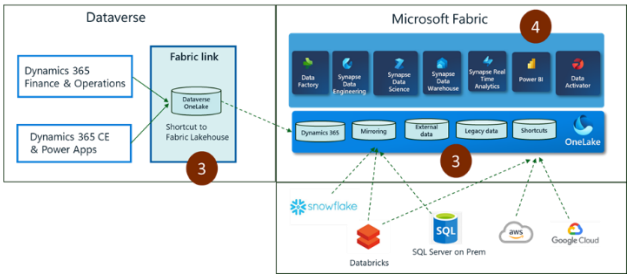
Example 1: Transition from “BYOD” and “Export to Data lake” to Fabric Link

Consider the case where you transition to Fabric Link from Export to Data lake.

Before – Export to Data lake



After – Fabric Link



As indicated in the before and after diagrams above, customer retired Export to Data lake service (1) as well as staging data stores (2) with Fabric Link. For operational insights, (4), they were able to consume data in OneLake directly in Power BI. Some of the insights require data merge, transformation and aggregation -(3). Instead of using disparate Azure services, they standardized on same tools built into Microsoft Fabric.

As we discuss below, innovations in Dataverse and Microsoft Fabric enable simplifications and cost reductions.

1. Dataverse comes with a built-in OneLake store. Operational data from Dynamics 365 and PowerApps are replicated to built-in lake store near real-time (to avoid impact to operational workloads) and linked securely to Fabric via shortcuts. There is no need to bring Azure storage and secure data that's exported out. Your data doesn't leave Dataverse governance boundary and authorized users in Fabric can work with data using all Fabric workloads.
2. Export to Data lake (as well as DES) exports data in CSV format. CSV files are not suited for direct consumption due to poor query performance as well as occasional read/write contention issues. "Before solution" uses Azure Data Factory to periodically ingest and convert raw data into a SQL Azure DB or an Azure data warehouse. This layer is not needed in the "after solution" since Dataverse built-in OneLake data is maintained in Dela/ parquet format – the same open format that is native to Microsoft Fabric. Delta/ parquet format, along with optimizations in Fabric removes the need to maintain additional data stores for caching and improving query performance while eliminating read/ write contention. You can create Power BI DirectLake reports directly over data in OneLake without any additional data stages.
3. While operational insights can be performed using the data already available in OneLake, you may have additional data from other systems. This data may need to be combined, reshaped and aggregated with Dataverse data. Microsoft Fabric provides an integrated environment which provides best of breed tools like Data flows, Data factory, Spark. As opposed to configuring and provisioning tools standalone, you can simply consume the tools of choice. Integrated billing, source control and security enables simpler management and governance.
4. While you can continue to use Power BI service for reporting purposes, Fabric introduces DirectLake mode reporting which leverages the in-memory indexes built into delta/ parquet format thereby removing the need to use Power BI import mode reports.

These innovations yield end-to-end cost savings in addition to the benefits discussed above. Following tables outlines the Line items of costs along with a comparison of before and after solutions. You can use the table below as a guideline to estimate expected cost savings.

Category / Line item	Before cost - **Export to data lake**	After cost - Fabric Link
Set-up & configuration	Need to use multiple tools - Pay for multiple software licenses/ subscriptions - System configuration efforts - effort to build/ validate data pipelines - continued governance, management, and monitoring - training of users	Simpler configuration experience in Power platform. Purchase Fabric capacity and pay for use for all services. You only pay for what you use – in some cases, you are billed by the second[1]. No ETL pipelines needed for Dataverse data
Data Staging (1, 2)	Cost incurred for Azure services - Azure storage cost including cost of IO - Synapse Analytics (SQL serverless query) - Data Factory jobs to copy data - Staging data stores (ex. SQL DB)	Cost increase in Dataverse DB storage Ex. if you sync 500GB of data from D365, Dataverse storage may increase by ~100GB (5~8x data compression)
Operating Costs - Data Prep (3)	- Azure storage staging area - Data pipelines - Data ingestion into SQL or Dwh - Development/ maintenance Data	Spend shifts to Fabric where you pay for consumption with a shared capacity.
Operating Costs – Reporting (4) - Power BI datasets - Reporting	Synapse analytics (SQL Query) Power BI capacity and storage for import mode reports	As your data is compressed, (ex. 1/3 ~ 1/6 original size) your reporting and query costs reduce accordingly. New features like DirectLake reporting reduces consumption of Power BI storage consumption.

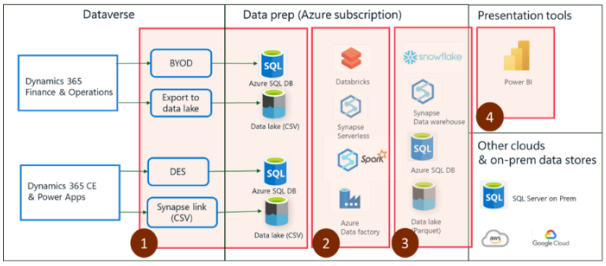
Example 2: upgrade to Synapse Link

Upgrading to Synapse Link is an option to consider if you are not currently using Fabric – or not planning to transition in the coming months.

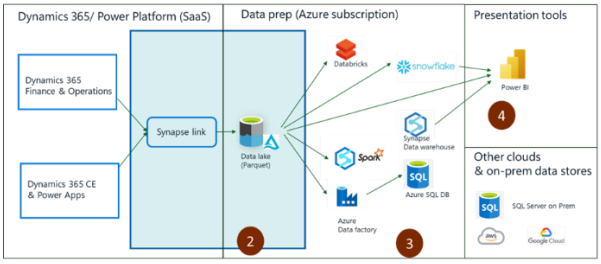
As indicated in the before and after diagrams below, customer retired Export to Data lake service (1) as well as staging data stores (2) with Synapse Link. Data stores used for data staging such as intermediate Azure SQL DBs are retired as Dataverse data is saved in Delta parquet format – a more performant and compressed data format. Downstream pipelines used for data merge, transformation and aggregation, (3), as well as (4) presentation tools such as Power BI remain unchanged.

[1] To do: Fabric pricing link

Before – Export to Data lake



After – Synapse Link



Synapse Link provides an easy upgrade path for customers looking to extend their existing investments with minimal changes. As we discuss below, innovations in Dataverse enable easy transition from FnO services like Export to Data lake and BYOD.

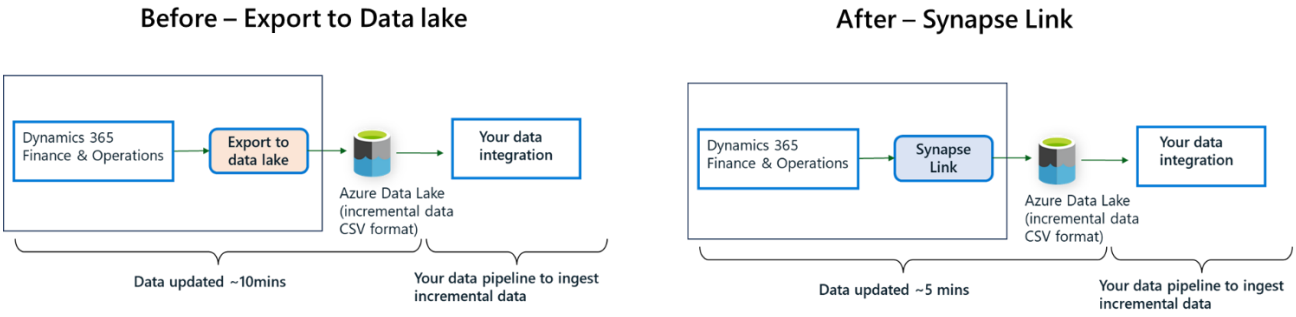
1. Synapse Link enables easy configuration of Azure storage and Synapse services within the Power platform maker portal. Several limitations of Export to Data lake are removed
    a. A uniform experience enables choosing data from all Dynamics 365 including FnO as well as Dynamics 365 CE apps. You can maintain a single service to export data from all Dynamics and PowerApps.
    b. Table limitations present with Export to Data lake go away, you can choose upto 1000 tables for each Synapse Link profile within a single storage account. You can create multiple profiles (ie. export pipelines) each with up to 1000 tables if desired.
    c. Synapse Link provides built-in support for using firewall restricted storage accounts to export data.
    d. No need to configure Azure resources, Synapse Link provisions and configures Azure resources on your behalf
    e. Synapse Link exports data in the same format as Export to Data lake or BYOD enabling you to retain existing pipelines
2. Export to Data lake (as well as DES) exports data in CSV format. CSV files are not suited for direct consumption due to poor query performance as well as occasional read/write contention issues. “Before solution” uses Azure Data Factory to periodically ingest and convert raw data into a SQL Azure DB or an Azure data warehouse. This layer is not needed in the “after solution” since Synapse Link exports data in Delata/ parquet which removes the need to maintain additional data stores for caching and improving query performance while eliminating read/ write contention.
3. You can continue to use your existing data pipelines for combining, reshaping and aggregating additional data from other systems.
4. You can continue to use Power BI service for reporting purposes, Fabric introduces DirectLake mode reporting which leverages the in-memory indexes built into delta/ parquet format thereby removing the need to use Power BI import mode reports.

These innovations yield end-to-end cost savings in addition to the benefits discussed above. Following tables outlines the Line items of costs along with a comparison of before and after solutions. You can use the table below as a guideline to estimate expected cost savings.

Category / Line item	Before cost - **Export to data lake**	After cost – **Synapse Link**
Set-up & configuration	Need to use multiple tools - Pay for multiple software licenses/ subscriptions - System configuration efforts - effort to build/ validate data pipelines - continued governance, management, and monitoring - training of users	Simpler configuration experience in Power platform. No ETL pipelines needed for Dataverse data
Data Staging (1, 2)	Cost incurred for Azure services - Azure storage cost including cost of IO - Synapse Analytics (SQL serverless query) - Data Factory jobs to copy data - Staging data stores (ex. SQL DB)	Synapse Link requires you to provide a spark pool to convert data to parquet format. Depending on the frequency of data sync, as well as the volume of data changes spark pool costs may vary. Ex. small/ medium data changes. - $600~2k for Hrly - $1,200~4,100 for 15min Ex. for Medium/ large data changes - $1,200~$2,500 for Hrly - @2,500~$8,300 for 15min
Operating Costs - Data Prep (3)	- Azure storage staging area - Data pipelines - Data ingestion into SQL or Dwh - Development/ maintenance Data	Same costs as before – however, you may not need to aggregate Dataverse data due to parquet conversion.
Operating Costs – Reporting (4) - Power BI datasets - Reporting	Synapse analytics (SQL Query) Power BI capacity	As your data is compressed, (ex. 1/3 ~ 1/6 original size) your reporting and query costs reduce accordingly.

Example 3: incrementally ingesting data to a data warehouse

If you are currently consuming incremental data from Export to Data lake feature to populate a downstream data pipeline, you can continue to use the same pipeline. As shown below, Synapse Link service can export incremental data changes in the same format as Change feeds in Export to Data lake.



Synapse Link service provides several enhancements over Export to Data lake for incremental data changes.

- Initial data load is included within change folders. This makes it easy for the same pipeline to consume both the initial load as well as incremental updates.
- Change data is not deleted in case of a re-initialization of a table.
- You can configure how often you want the change feeds updated in Synapse Link – with as low as 5mins.
- System creates a time-based folder structure and metadata that helps you read the changes in chronological order. This approach makes is more economical to read only the changes using common Data ingestion tools like Azure data factory.

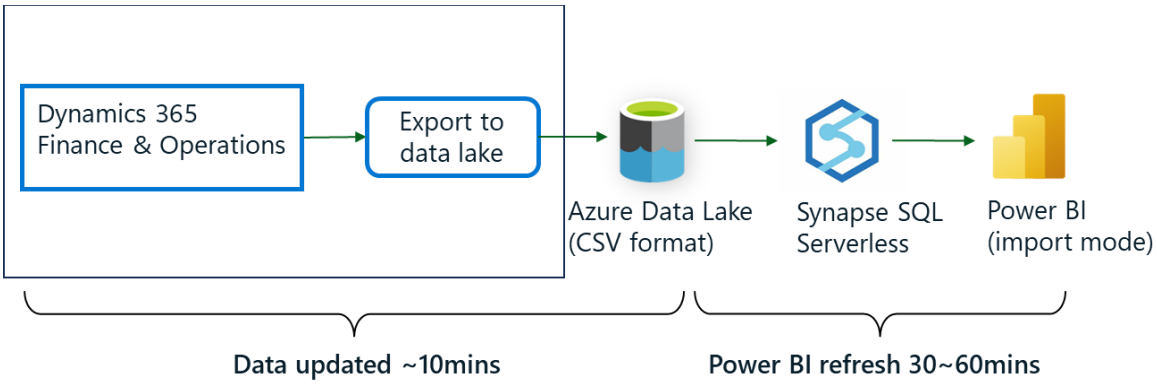
Before and after cost changes are minimal in this case.

Understanding benefits – more real-time reporting

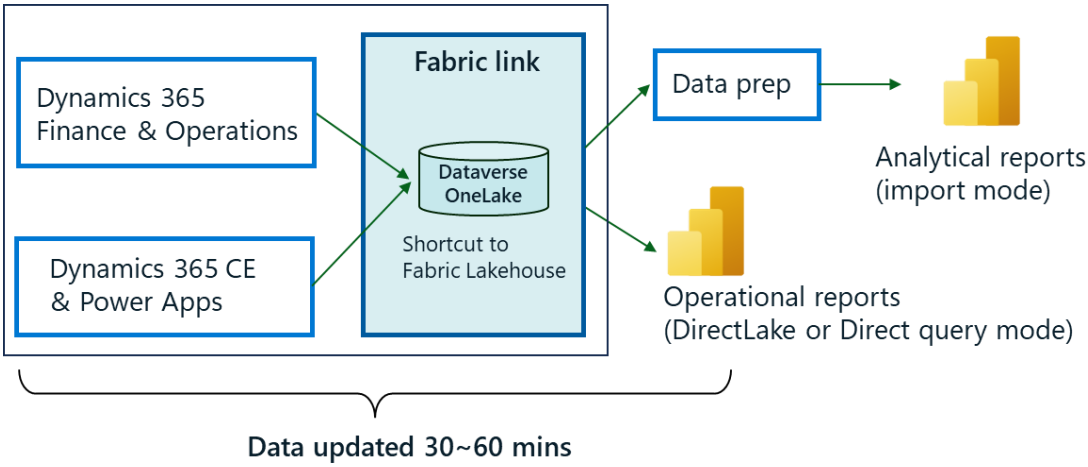
Microsoft Fabric not only simplifies your data integration architecture, but also reduces the need to copy or replicate data. The underlying data format in Fabric, the industry standard Delta parquet format, can be directly consumed by Power BI, Data flows, notebooks and other workloads eliminating the need for caching and staging areas.

If you upgrade to Synapse Link and continue to export data, your data is saved in Delta/ parquet format which may reduce the steps in your own data pipelines for operational reporting.

Consider the time taken to refresh an operational report such as Inventory analysis or month-end Financial analysis. These reports may require data aggregation of millions of rows of data from multiple tables in Dynamics 365 Finance. Using export to data lake service, CSV data is exported within 10mins. This data may need to be imported into a Power BI report to provide better response times. Power BI refresh can be performed up to 24 times/ day (ie. every 30mins) and depending on how the report is designed, refresh may take several minutes to complete refresh. Using this approach, users can see data within 60mins of an update.



Operational reports that source data using Fabric Link can leverage Direct Lake mode or DirectQuery mode in Fabric which leverages the in-memory index built into Delta/parquet files. In these modes, you don't need to schedule refresh of Power BI reports as the report always shows latest Dataverse data updated in Fabric.



Fabric Link service updates data in Dataverse OneLake within the hour as of this point in time. Dataverse triggers data update jobs every 15mins and depending on the volume of data changes, you may see updated parquet files within 30~60mins. We are working on reducing this frequency in the coming months to enable more real-time reporting.

If you are consuming incremental data feeds from Dynamics 365 with export to data lake service for near-real time data integration scenarios (ie. example 3), upgrade to Synapse link will enable you to run the same data pipelines.

Known issues and workarounds

Currently, there are several limitations that are being addressed by the product team. You can use a suggested workaround until we address these limitations. To learn more about the upcoming roadmap and stay in touch with the product team, join the preview Viva Engage group.

Known issue	Workaround	Fix and roadmap
When adding a large number of tables at once, the system makes an initial copy of data. **We have seen rare cases (especially in smaller environments and Tier 2 sandboxes) where operational workloads may slow down and Initialize time may become much longer**	We are addressing this issue which typically impacts smaller environments. Add 5 tables at a time in case your environment is a Tier-2 Sandbox – once the initialization completes, you can add more.	Apr release resolves some of these issues We are redacting varBinary fields and varBinary attachments from tables added to Synapse Link and Fabric Link. These fields will be ignored. • PU 63 cumulative update 7.0.7198.95 • PU 62 cumulative update 7.0.7120.155 Fabric Link feature scales initialization workloads up and down as resources permit at roughly 2 concurrent tables per AOS. Ex. if you have 5 AOS servers in your environment, system concurrently initializes up to 10 tables.
When adding a large number of tables at once, the system makes an initial copy of data. In some cases, especially with very large tables, initialization may take longer. **In some cases initialization appears to be stuck for several days.**		Apr release introduces 2 features to reduce this issue. Faster initialization of large tables (>200m rows) • PU 63 cumulative update 7.0.7198.91 • PU 62 cumulative update 7.0.7120.152 We have enabled indexes to enable faster data sync. In case there's an ongoing transaction in the operational database, index creation needs to wait for completion of the transaction. This wait will delay initialization process. System detects such index delays and informs the user to take action if the delays are excessive (ex. > 24hrs).
In case your Dataverse environment is located in an Azure region different than the one where your Fabric capacity is located, you can't use the Link to Fabric feature	This issue is addressed in Apr release	In Apr release, you can Link to a Fabric capacity located within the same Geo boundary (ex. USA) NOTE: you may incur networking charges in Fabric. See:
In case your Dataverse environment is	You need to create a data lake	In Apr release, you can link to Storage

located in an Azure region different than the one where your Data lake or Synapse workspace is located, you can't use the Synapse Link feature	and synapse resources in the same region as Dataverse	account located within the same geo boundary. NOTE: you may incur networking charges in Azure resources like Data lakes. See:
AOS authorization is a way to secure sensitive data fields in FnO against data exfiltration scenarios. **If the table selected contains data columns that are secured via AOS Authorization, those columns are ignored and the exported data doesn't contain the column.** **Ex *CustTable*,** **column *TaxLicenseNum* has the metadata property AOS Authorization set to Yes. This column is ignored**		Apr release addresses this issue With FnO latest quality update (10.0.38, 10.0.39), AOS authorization fields will be added - Incremental updates will include this column - Modified records will show these columns and value - Full refresh will include these fields and all values
If the table selected contains data columns that are of Array type, those columns are ignored and the exported data doesn't contain the column. **For example, in a custom table named *WHSInventTable*,** **columns FilterCode and FilterGroup are of type array. These columns aren't exported with Azure Synapse Link**	There is no workaround to this issue. We are working on enabling these fields.	ETA – May update
In case of finance and operations app tables that exhibit valid time stamp behavior, only the data rows that are currently valid are exported with Azure Synapse Link. For example, **the ExchangeRate table contains both current and previous exchange rates. Only currently valid exchange rates are exported in Azure Synapse Link.**	We are working on enabling this pattern so that you can extract all records. As a workaround, until this issue is fixed, you can use an Entity such as **ExchangeRateBIEntity**.	ETA – May update
Export more than 1000 tables	We are enabling you to select more than 1000 tables in Fabric Link and in a Synapse Link profile If you are using Synapse Link, you can workaround by creating 2 or more profiles that contain less than 1000 tables	ETA – May update

Table inheritance and derived tables are concepts in finance and operations apps. When choosing a derived table from finance and operations apps, fields from the corresponding base table currently aren't included.	You need to select the base table in addition to the derived table if you need access to these fields. You can use the FastTrack solution provided via GitHub[2] - this solution creates view(s) which include fields from base tables	ETA – Jun update
Finance and operations apps tables included in an Azure Synapse Link profile can't be migrated to a different environment using the import and export profile feature in Azure Synapse Link	We are working on this feature	ETA – Jun/ July
Synapse Link or Fabric Link enables Entities where "change tracking" property is enabled. **Currently, change tracking can't be enabled for all finance and operations entities. The Track changes checkbox is unavailable for entities created in finance and operations in the past for data migration.** **In some entities, enabling change tracking might fail with the error message "chosen entity doesn't pass the validation rules..." or the Track changes checkbox is disabled for some entities.** **For more information about entity validation rules and how you can fix them, go to Enable row version change tracking for data entities. You might need developer assistance to complete the steps.**	If the chosen entity is unavailable because of the change tracking limitation, you can choose the tables that comprise the data from that entity. You can use EntityUitl[3] solution provided by the FastTrack team to create Entity shapes using tables	ETA – Jun/ July

[2] https://github.com/microsoft/Dynamics-365-FastTrack-Implementation-Assets/tree/master/Analytics/DataverseLink/DataIntegration#derived-tables

[3] https://github.com/microsoft/Dynamics-365-FastTrack-Implementation-Assets/tree/master/Analytics/DataverseLink/DataIntegration/EntityUtil

In case of a database restore operation in Dataverse, finance and operations entities enabled in Azure Synapse Link are removed.	To re-enable entities, you need to re-enable corresponding virtual tables for all selected entities, re-enable change tracking, and reselect the tables in Azure Synapse Link	ETA: Aug/ Sept
Finance and operations apps tables added to an Azure Synapse Link profile might be removed when a back-up is restored in Dataverse.	You must add finance and operations tables into the profile after a database restore operation. Go to Known limitations with finance and operations tables for details on re-enabling entities after a database restore operation.	ETA: Aug / Sept

Links to learn more

You can visit product documentation for more information on configuring Fabric Link or Synapse Link
<< links>>

You can visit TechTalks for detailed steps and demos for transitioning.
https://aka.ms/TransitiontoSynapseLinkVideos

You can join the preview Viva Engage group to stay in touch with the product team as well as thousands of customers and partners. In addition to an active discussion forum, there are weekly office hours with demos and provisions to ask questions in person. This group requires sign-up at https://aka.ms/SynapseLinkforDynamics