SPSync

ADF vs. replication decision guide (PDF)

Choosing between a Data Factory pipeline, a Fabric dataflow or Copy job, a Power Automate flow and a replica, one question at a time.

Finance wants the Projects list in SQL by the end of the month, and the data team has four Microsoft-shaped answers on the table. Each works for some lists. The trouble starts two years later, when the answer picked for a small plain list is still running against a large one with lookups, and the person who built it has moved team. The Microsoft facts below come from Microsoft Learn, checked on 20 October 2026.

The four options for copying a SharePoint list

Azure Data Factory | Fabric Dataflow Gen2 or Copy job | Power Automate | A replica kept by SPSync

What it is | A pipeline with a Copy activity per list | Power Query or a Copy job in a Fabric workspace | Actions per item or on a schedule | Read-only replication of each list into a database

How it is billed | Pipeline activity and integration runtime usage | Fabric capacity units | Per user or per flow licence | Licence per installation and per target family, annual terms

Microsoft's Fabric connector page lists Dataflow Gen2, a pipeline Copy activity with the Lookup activity, and a Copy job, all as sources for a SharePoint Online list. If you read somewhere that Fabric pipelines cannot read SharePoint lists, that is out of date. The Copy job entry says full load only, so every run reads the whole list again.

Lookups and person columns in the copy

In Data Factory's documented type mapping, a lookup column and a Person or Group column both arrive as Edm.Int32. That is the target item's id with no display value, so something downstream has to resolve it. Managed metadata and attachments are listed as unsupported. SPSync copies a managed metadata column as the term's GUID and label, and does not copy attachments. Dataflow Gen2 goes through Power Query, whose v2.0 SharePoint connector is limited to 12 join operations per query, so a list with many lookup and person columns needs a trimmed default view before it will load.

A flow maps fields one action at a time. SPSync writes each lookup as two columns, {col}_id and {col}_value, and each person as a key into one shared sp_users table. Multi-value columns become child tables. The schema reference has the full mapping.

Deleted items in SharePoint lists

Deletes are where copies drift. The Data Factory list connector does not document delete handling at all. A Fabric Copy job on a SharePoint list is full load only, so finding what was deleted means

comparing each full copy with the one before. Power Automate has a "When an item is deleted" trigger, and Microsoft's connector reference says only site collection admins of that site can use it.

SPSync keeps the row and sets sp_deleted_at, so reports can filter on it and an auditor can still find the item. The generated reporting views leave deleted rows out.

Large lists and the 5,000-item threshold

SharePoint blocks a query that would scan more than 5,000 items without an indexed filter, so a pipeline or flow paging the whole list is written around that. The SharePoint connector in Power Automate is also throttled at 600 API calls per connection every 60 seconds. SPSync reads with delta queries after the first load, so later runs fetch only what changed, and it honours Graph's Retry-After when SharePoint asks it to slow down.

Cost model for each option

Cost | Data Factory | Fabric | Power Automate | SPSync replica
Platform | Per activity run and integration runtime | Capacity units consumed per refresh | Licence per user or flow; the SQL Server connector is Premium | Annual licence, quoted after a call
Mapping work | Per column, by hand | Per query step | Per field, per flow | None; the schema is generated
Deletes | Design your own | Compare full copies yourself | Separate admin-only trigger | Soft delete by default

A short schedule on Fabric consumes capacity on every run, and a flow copying item by item costs more actions as the list grows. A replica's run cost follows what changed. Across all four, the licence is the smaller part of the bill once you count the hours.

Who owns the schedule

Each schedule needs an owner for its alerts and credentials. Data Factory usually sits with the data platform team, plus a service principal and certificate to rotate. Fabric sits with the workspace owner. Flows tend to belong to a site owner, which is how they end up unowned.

SPSync runs on a schedule you set, every five minutes by default, with a reconcile that re-checks every item weekly by default as QA and after an outage. On the command line you call spsync run from cron or a CI pipeline, so the schedule sits with your other jobs.

Picking one for your list

If this is true | Use
The list is one source among many already in a Data Factory estate | Data Factory
Every report already lives in Fabric and the list is small and plain | Fabric Dataflow Gen2
You need to react to one item changing, such as an approval | Power Automate
The copy has lookups, deletes or a large list, and people depend on it daily | A replica

Plenty of teams run more than one. Keep the flow that sends the approval email and move the

copying to a replica. Replacing Data Factory pipelines and scheduled flows covers the swap, and the comparisons with Azure Data Factory, Fabric dataflows and Power Automate go further on each tool.

Full, up to date version: https://spsync.io/resources/adf-vs-replication-decision-guide/