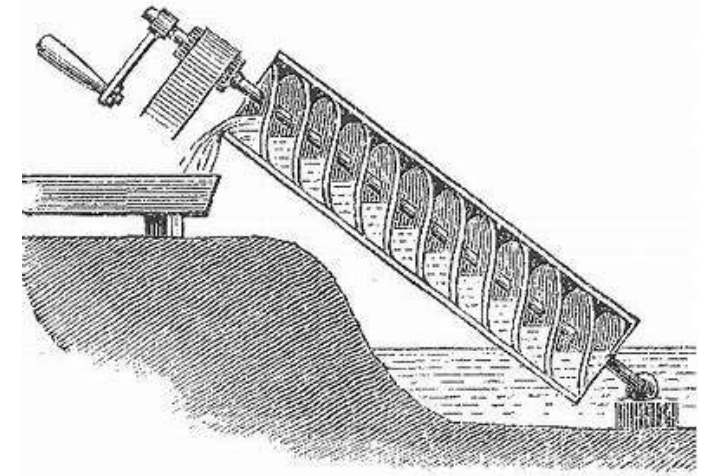


Microsoft Graph Data Connect for SharePoint

Delivering rich data assets, analytics and insights to SharePoint tenants



“Project Archimedes”



Microsoft Graph Data Connect for SharePoint

Problem Statement: SharePoint and OneDrive tenants are demanding more **analytics and insights** into how their organization use M365 products.

Current M365 Solutions:



Admin Portal – specific reports, top n lists



PowerShell – command-line, interactive



Graph API – custom applications



Third-party Software – ISV partner solutions



Microsoft Graph Data Connect for SharePoint

Problem Statement: SharePoint and OneDrive tenants are demanding more **analytics and insights** into how their organization use M365 products.

Current M365 Solutions:



Admin Portal – specific reports, top n lists



PowerShell – command-line, interactive



Graph API – custom applications



Third-party Software – ISV partner solutions

Securely deliver **rich data assets directly** to a tenant's Azure account, where they can run their own analytics.

- Big data: scales to billions of objects
- M365 Data: SharePoint, Teams, AAD, Exchange and more
- Joinable with company data
- No throttling. 1 request = 1 dataset
- Approval workflow per request
- One-time trigger or recurring
- Integrated with Azure data tools
- Microsoft-provided schema, samples and templates



Microsoft Graph Data Connect for SharePoint

The Problem

SharePoint and OneDrive tenants are demanding more **analytics and insights** into how their organization use M365 products.

The Solution

Deliver **rich data assets** directly to a customer's Azure subscription. Customers then build their own apps, reports and dashboards using tools like Azure Synapse and Power BI.

How is this different?

- Focuses on **enterprise** customers (security, compliance)
- Delivers entire "**data stories**" with one request
- Targets "**data people**", who will build with these datasets
- Works at **scale** (billions of data rows)

More information:

This uses the [Microsoft Graph Data Connect](https://aka.ms/SharePointData), which offers more datasets, including Exchange and Teams. Details and Setup instructions at <https://aka.ms/SharePointData>

| Dataset | Scenarios | | |
|--------------|-----------|----------|--------|
| | Security | Capacity | Health |
| Permissions | X | | |
| AAD Groups | X | | |
| SPO Groups | X | | |
| Sites | X | X | X |
| Files | X | X | |
| File Actions | | X | |
| Sync Health | | | X |
| Sync Errors | | | X |

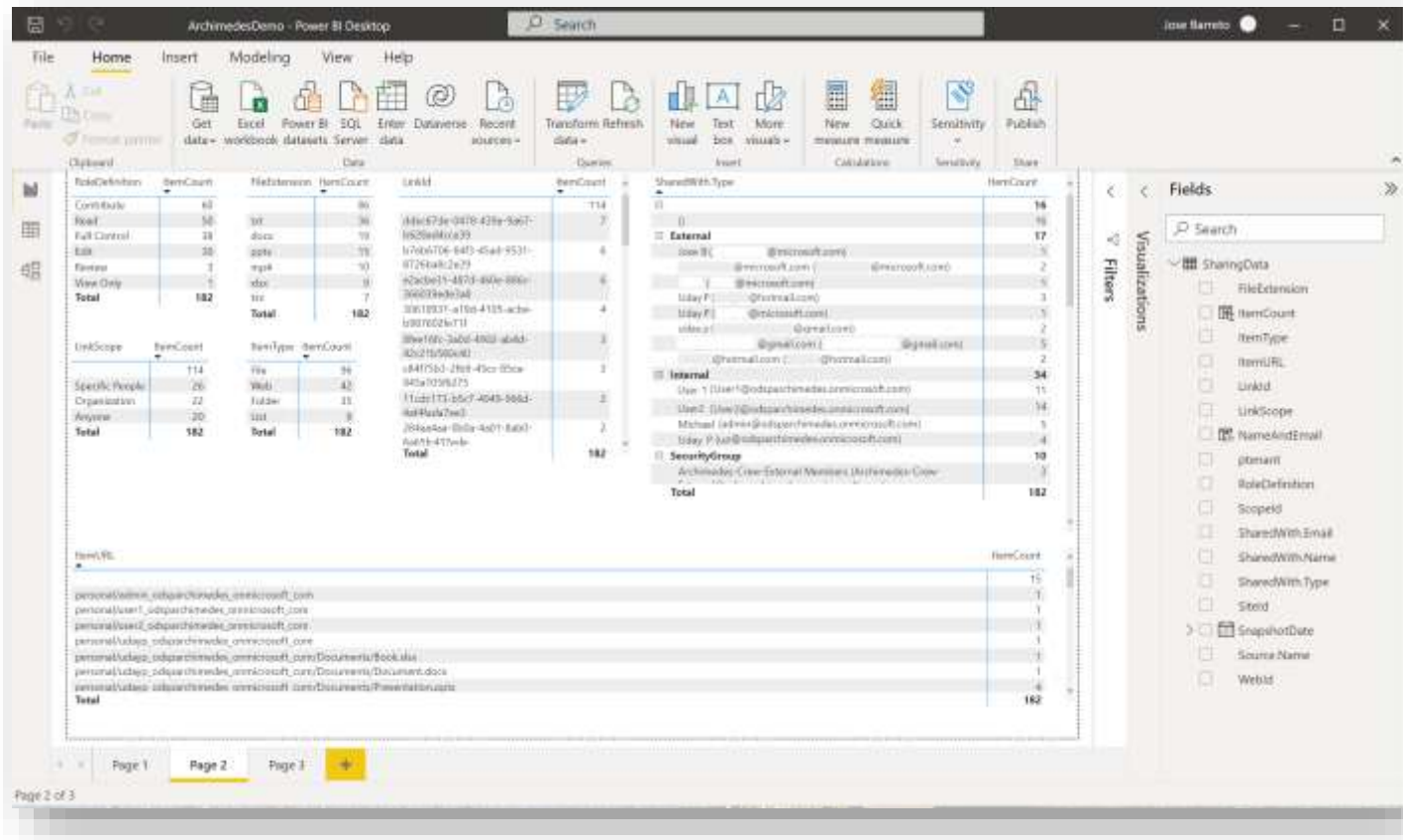


Microsoft Graph Data Connect **Datasets**

- Microsoft Entra ID
 - User
 - Manager
 - Direct Report
- Outlook and Exchange
 - Calendar View
 - Conference Room Calendar
 - Contact
 - Event
 - Inbox
 - Mailbox Setting
 - Mail Folder
 - Message
 - Sent Item
 - Todo Task Folder
 - Todo Task
 - Outlook Group Conversation
- Viva Insights
 - Person Report
- Microsoft Teams
 - Team Chat
 - Teams Call Records
 - Teams Channel Details
 - Teams Standard Channel Messages
 - Teams Transcript
- Microsoft Groups
 - Group Details
 - Group Members
 - Group Owners
- OneDrive and SharePoint
 - SharePoint Sites
 - SharePoint Groups
 - SharePoint Permissions
 - SharePoint Files
 - SharePoint File Actions
 - OneDrive Sync Health
 - OneDrive Sync Errors



Sample: Permissions Dataset

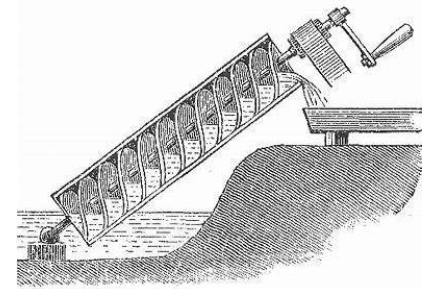
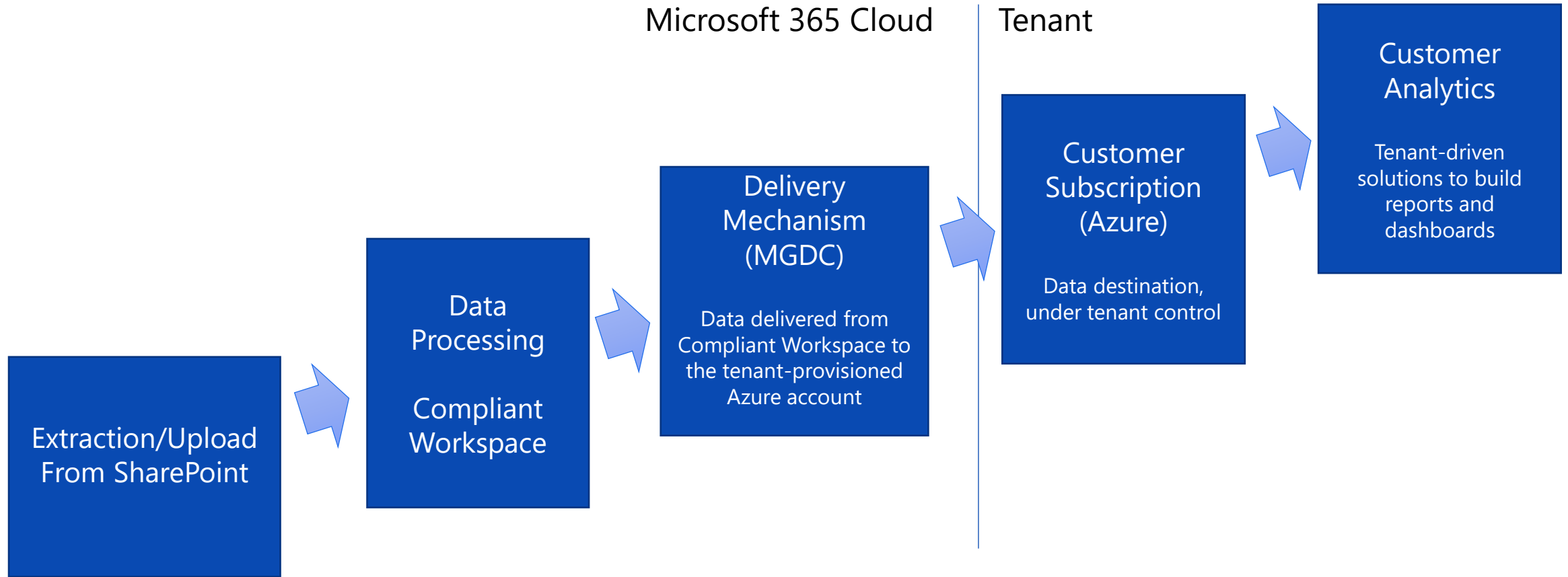


```

{
  "tenant": "3abcdefg-4d67-4c21-842a-abcea48840d5",
  "SiteId": "191f7f71-a950-432b-809f-09043cec1d88",
  "WebId": "0e57e46e-3215-4eb3-ab53-c9cc5a44a683",
  "ListId": "1027718d-17dc-4893-8ab4-83cea78fc3b6",
  "ListItemId": 2,
  "UniqueId": "487c5a4c-7cf0-4127-b3d8-7b8beda75d96",
  "ItemType": "File",
  "ItemURL": "sites/ProjectX/S2023/M_AllUp.json",
  "FileExtension": "json",
  "RoleDefinition": "Contribute",
  "LinkId": "b31c4d99-de3a-4d71-9d22-606397106737",
  "ScopeId": "1d637c41-5bb6-4d65-b9c2-99ddf1c2b3e8",
  "LinkScope": "Anyone",
  "SharedWithCount": [
    {
      "Type": "SecurityGroup",
      "Count": 1
    }
  ],
  "SharedWith": [
    {
      "Type": "SecurityGroup",
      "Name": "ProjectX1",
      "Email": "ProjectX1@contoso.onmicrosoft.com",
      "TypeV2": "SecurityGroup",
      "AADObjectId": "40cceb15-526d-4562-9172-fdc8f5be2001",
      "UserCount": 4
    }
  ],
  "ShareCreatedBy": {
    "Type": "User",
    "Name": "User3",
    "Email": "User3@contoso.onmicrosoft.com",
    "UPN": "user3@contoso.onmicrosoft.com"
  },
  "ShareCreatedTime": "2023-05-06T19:56:39Z",
  "ShareLastModifiedBy": {
    "Type": "User",
    "Name": "User3",
    "Email": "User3@contoso.onmicrosoft.com",
    "UPN": "user3@contoso.onmicrosoft.com"
  },
  "ShareLastModifiedTime": "2023-05-06T19:56:39Z",
  "TotalUserCount": 4,
  "Operation": "Full",
  "SnapshotDate": "2024-07-31T00:00:00Z"
}

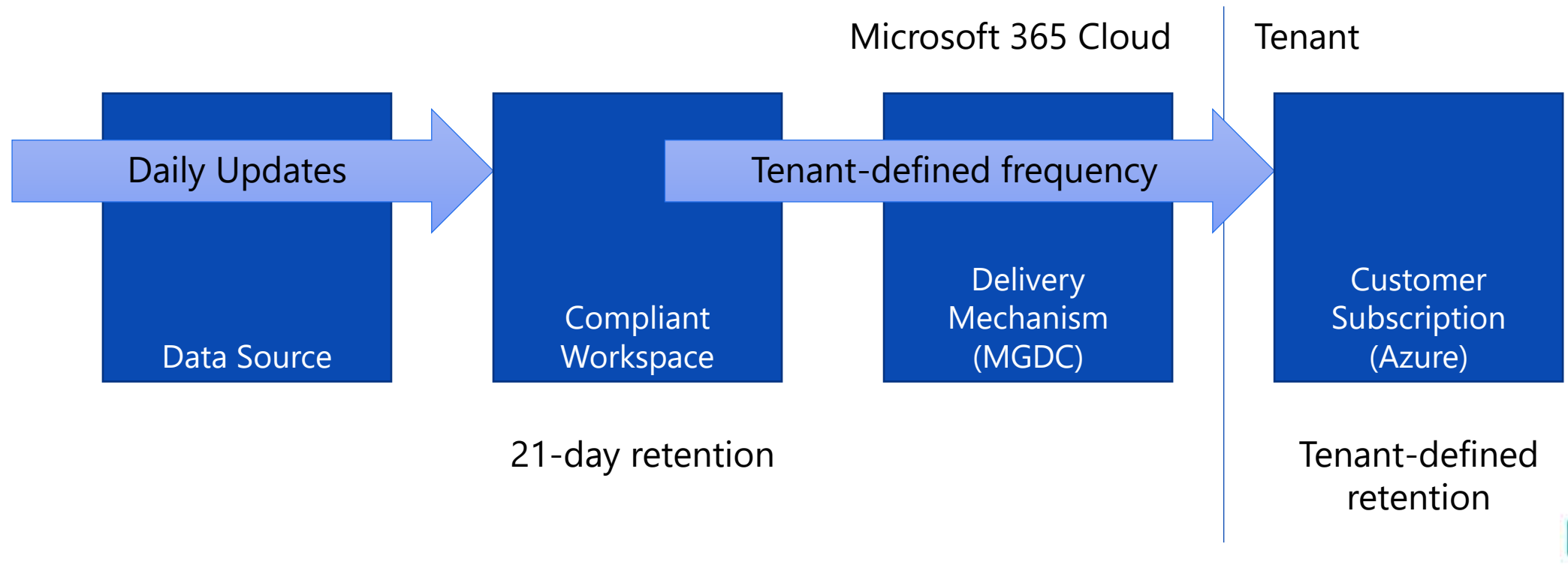
```

Steps to Process the Data



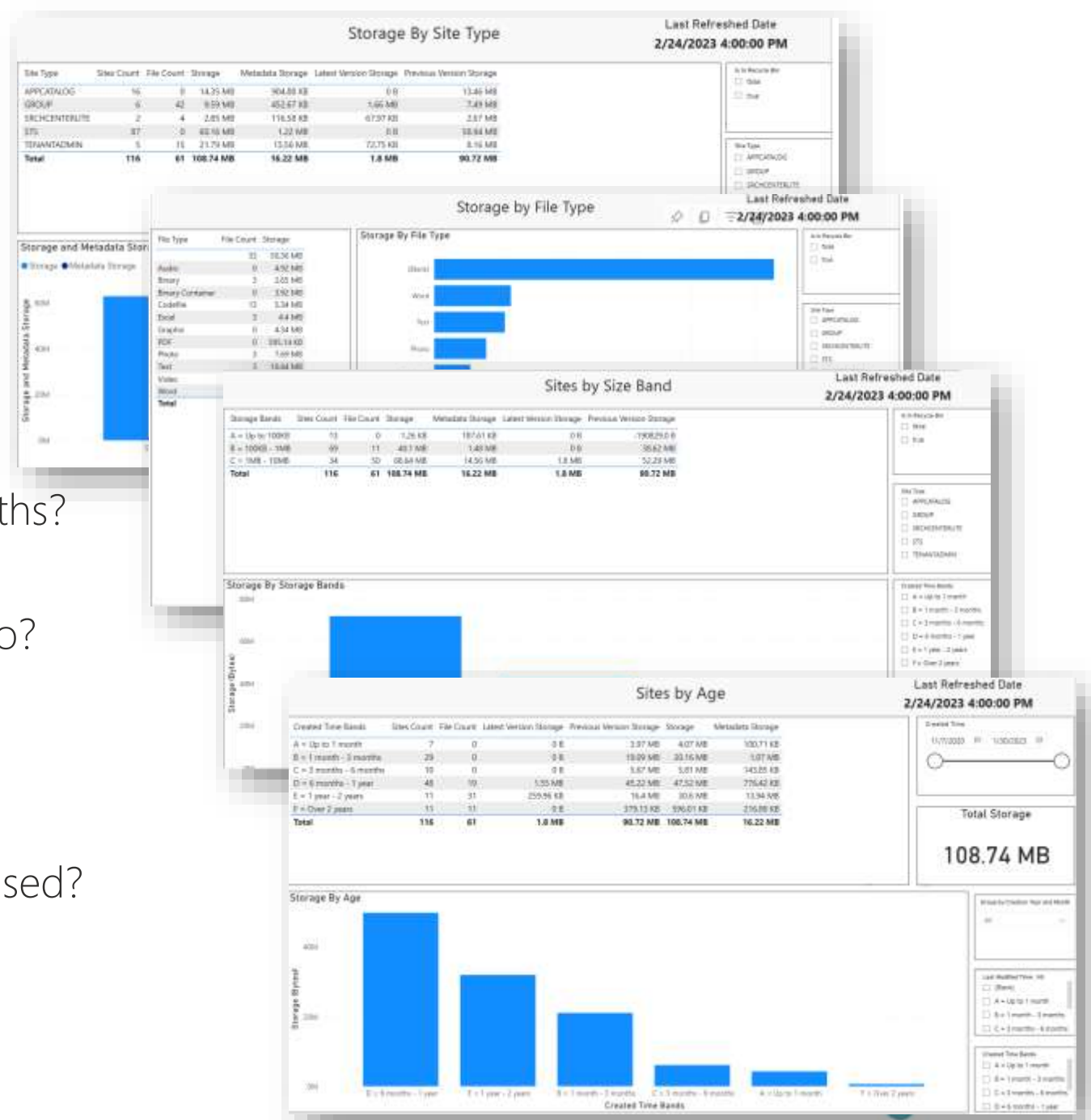
Updates and Retention

- Data updated daily, at off-peak hours for each region
- Data retained in the compliant workspace for 21 days
- Tenant decides retention of the data they pull into their Azure subscription



Scenario: Capacity

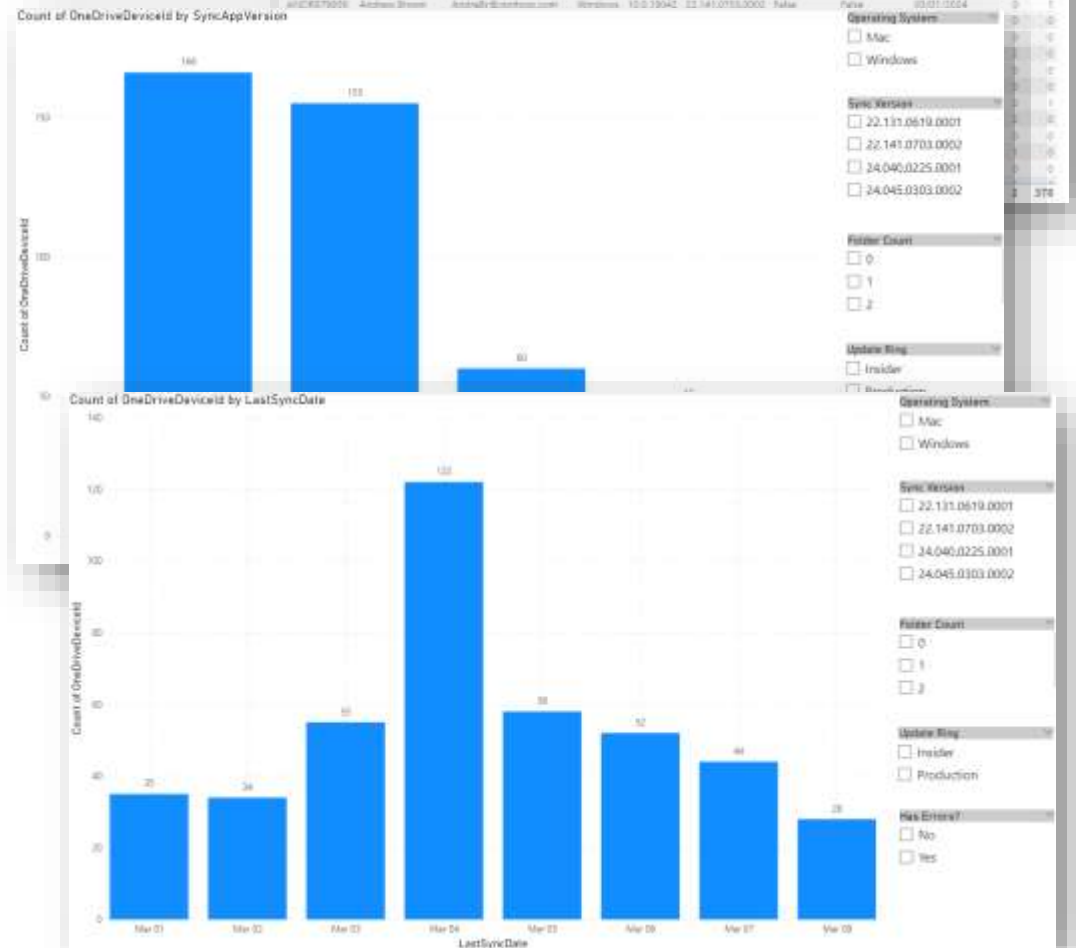
- What SharePoint sites are the largest?
- What type of site uses the most storage?
- What's the current storage for sensitive sites?
- How much is used by previous versions?
- Which sites were updated in the last few months?
- Which sites have just one owner?
- How many sites were created over 2 years ago?
- How many sites haven't changed in 1 year?
- How many sites have over 1TB of files?
- Which file extensions are most used?
- Are certain "bad" extensions (exe/pst) being used?
- Which file types use the most storage?
- How many files have more than 1GB?
- ...Or a combination of these questions



Scenario: Sync Health

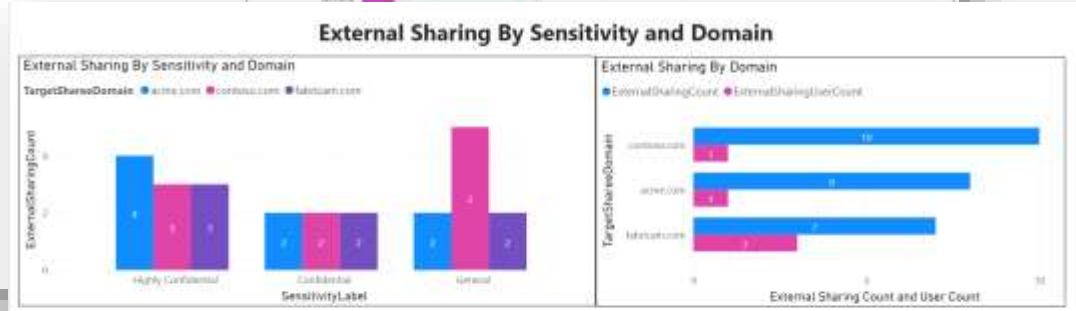
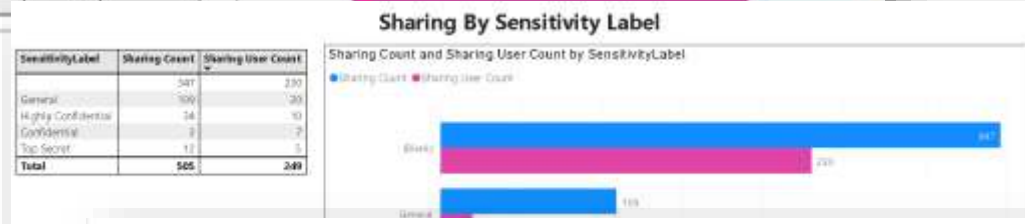
- How many devices are healthy?
- How many devices have opted in for Folder Backup?
- Which Folders are most selected for Folder Backup?
- What is the breakdown of unhealthy devices version?
- Which errors are making the devices unhealthy?
- What errors are only showing in a specific version?
- Which devices are showing a specific error?
- ...Or a combination of these questions

| DeviceName | UserName | UserEmail | OSName | OSVersion | SyncAppVersion | Alert Cpt. n | Wired Def'n | LastSyncDate | Folder | Errors |
|------------|------------------|-----------------------|--------|------------|------------------|--------------|-------------|--------------|--------|--------|
| ALAC137138 | Adrian Brown | adrianb@brown.com | Win | 11.6.1 | 22.131.0619.0001 | False | False | 03/03/2024 | 0 | 0 |
| AMC11447 | Aditya Reddy | adityar@reddy.com | Win | 10.0.19042 | 24.045.0303.0002 | False | False | 03/03/2024 | 1 | 1 |
| AD-0491823 | Adigal Pulurajan | adigalp@pulurajan.com | Mac | 11.6.1 | 22.141.0703.0001 | False | False | 03/03/2024 | 3 | 3 |
| AMRA10998 | Adityan Thiruk | adityat@thiruk.com | Mac | 11.6.1 | 22.141.0703.0002 | False | False | 03/03/2024 | 3 | 3 |
| ACMNS12901 | Adrian Cordar | adrianc@cordar.com | Win | 10.0.19042 | 22.131.0619.0001 | False | False | 03/01/2024 | 3 | 2 |
| ACR103089 | Adrian Horler | adrianh@horler.com | Win | 10.0.19042 | 22.131.0619.0001 | False | False | 03/01/2024 | 3 | 1 |
| ACR1114833 | Adrian Hooding | adrianh@hooding.com | Win | 10.0.19042 | 22.131.0619.0001 | False | False | 02/04/2024 | 1 | 0 |
| ACR102181 | Adrian Lee | adrianl@lee.com | Win | 10.0.19042 | 22.131.0619.0001 | False | False | 03/03/2024 | 1 | 0 |
| ALAV171318 | Alan Payne | alanp@payne.com | Win | 10.0.19042 | 22.141.0703.0002 | False | False | 03/04/2024 | 2 | 0 |
| ALAN823 | Alan White | alanw@white.com | Win | 10.0.19042 | 22.131.0619.0001 | False | False | 03/04/2024 | 1 | 0 |
| ALAV12353 | Alan Stevens | alanst@stevens.com | Win | 10.0.19042 | 22.141.0703.0002 | False | False | 03/03/2024 | 1 | 0 |
| ALSC40981 | Alan Chandler | alanch@chandler.com | Mac | 11.6.1 | 22.141.0703.0002 | False | False | 03/04/2024 | 2 | 0 |
| ALC131984 | Alan Taylor | alant@taylor.com | Win | 10.0.19042 | 22.131.0619.0001 | False | False | 03/04/2024 | 2 | 0 |
| ALAN6706 | Alan Smith | alans@smith.com | Win | 10.0.19042 | 22.141.0703.0002 | False | False | 03/04/2024 | 1 | 0 |
| ALLA12333 | Alan Young | alany@young.com | Mac | 11.6.1 | 24.045.0303.0001 | False | False | 03/03/2024 | 3 | 3 |
| AWD104808 | Angela White | angew@white.com | Win | 10.0.19042 | 22.131.0619.0001 | False | False | 03/04/2024 | 2 | 2 |
| AWD145370 | Andrew Mason | andrewm@mason.com | Mac | 11.6.1 | 22.141.0703.0002 | False | False | 03/03/2024 | 3 | 3 |
| AWD123332 | Andrew Smith | andrews@smith.com | Win | 10.0.19042 | 22.141.0703.0002 | False | False | 03/03/2024 | 1 | 2 |
| AWD131170 | Andrew Smith | andrews@smith.com | Win | 10.0.19042 | 22.131.0619.0001 | False | False | 03/03/2024 | 1 | 0 |
| AWD137188 | Andrew Hall | andrewh@hall.com | Win | 10.0.19042 | 24.045.0303.0001 | False | False | 03/03/2024 | 3 | 3 |
| AWD137900 | Andrew Brown | andrewb@brown.com | Win | 10.0.19042 | 22.141.0703.0002 | False | False | 03/03/2024 | 3 | 3 |

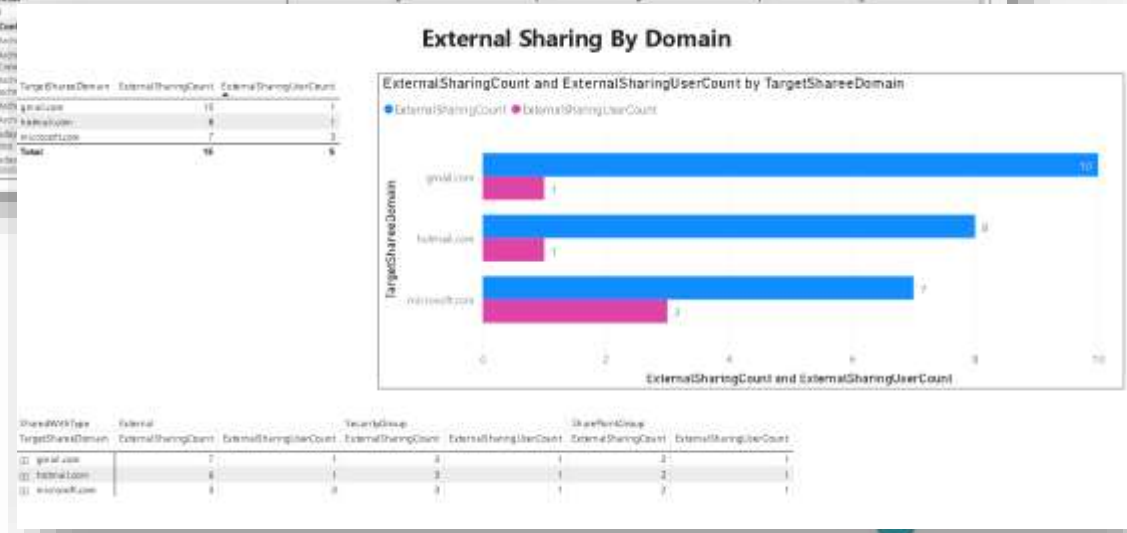


Scenario: Security

- Is oversharing happening?
- Is external sharing happening?
- Is sensitive data being shared?
- How much sharing per sensitivity label?
- Is sensitive data shared with external users?
- What external domains are being shared with?
- Which sites have been shared the most?
- What roles (levels of sharing) are being used?
- What permissions does a specific user have?
- What file extensions are most shared?
- How much sharing happens at the Web, Folder, List or File level?
- ...Or a combination of these questions



| TargetShareDomain | ExternalSharingCount | ExternalSharingUserCount | corpnet.com | ExternalSharingCount | ExternalSharingUserCount | azure.com | ExternalSharingCount | ExternalSharingUserCount | fabrikam.com | ExternalSharingCount | ExternalSharingUserCount |
|---------------------|----------------------|--------------------------|-------------|----------------------|--------------------------|-----------|----------------------|--------------------------|--------------|----------------------|--------------------------|
| Confidential | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| General | | | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Highly Confidential | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| System:FullControl | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| System:ListAndView | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| System:FullControl | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Contribute | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| System:Contribute | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| System:Read | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| System:ViewOnly | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| System:Review | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 16 | 9 | | | | | | | | | |






| ShareType | External | System:FullControl | System:ListAndView | System:Review | System:ViewOnly | |
|-------------------|----------------------|--------------------------|----------------------|--------------------------|----------------------|--------------------------|
| TargetShareDomain | ExternalSharingCount | ExternalSharingUserCount | ExternalSharingCount | ExternalSharingUserCount | ExternalSharingCount | ExternalSharingUserCount |
| gmail.com | 7 | 1 | 0 | 0 | 2 | 1 |
| fabrikam.com | 6 | 1 | 2 | 1 | 2 | 1 |
| corpnet.com | 3 | 0 | 0 | 1 | 2 | 1 |

Capacity Template in the Azure Synapse Gallery

Gallery

Database templates Datasets Notebooks SQL scripts **Pipelines**

capacity Category: All Tags: All

(x)   

Unlock Capacity Analytics and Insights for SharePoint and OneDrive using Microsoft 365 Datasets
by Microsoft

Use this template to accelerate capacity scenarios by understanding the storage and site lifecycle for S...

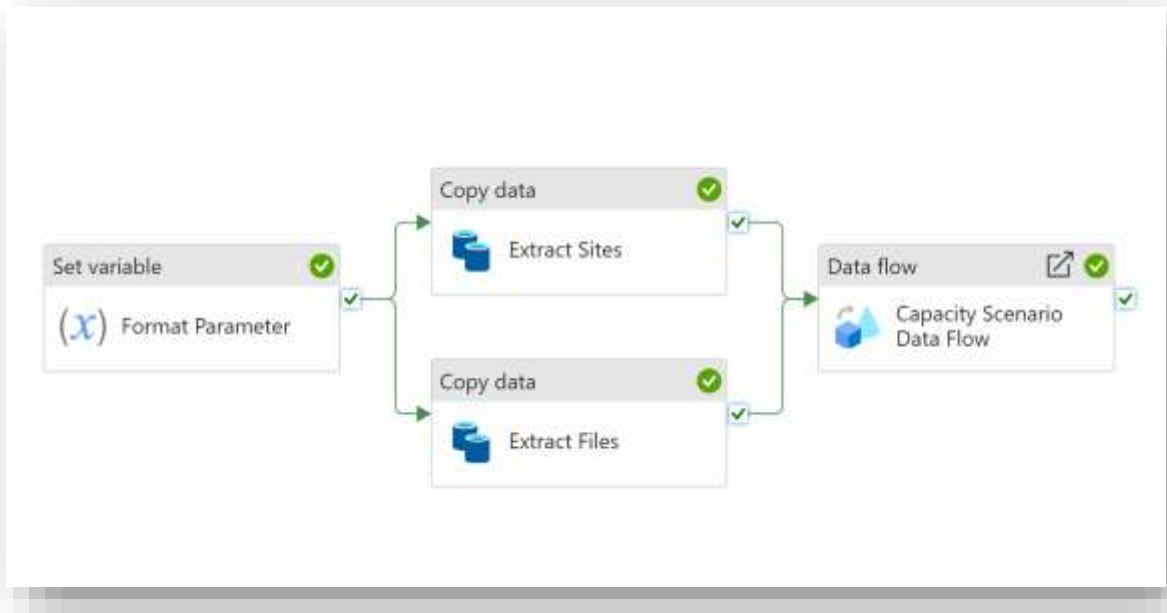
Description

Use this template to accelerate capacity scenarios by understanding the storage and site lifecycle for SharePoint and OneDrive in your organization. This template extracts Microsoft 365 (Office) SharePoint data via Microsoft Graph Data Connect to produce analytics-ready data for analysis.

We would love to hear your thoughts on this template. Please send us your ideas and feedback at [link].

Tags

- MGDC
- Azure Synapse Analytics
- OneDrive
- SharePoint
- Capacity
- Sites
- Files
- Documents
- Syntex
- M365
- Office365
- Graph
- Microsoft
- Office



[Capacity Template Setup](#)



Oversharing Template in the Azure Synapse Gallery

Gallery

Database templates Datasets Notebooks SQL scripts Pipelines

SharePoint Category: All Tags: All

Unlock advanced analytics and insights using Microsoft 365 SharePoint datasets by Microsoft

Use this template to accelerate sharing security scenarios by identifying information sharing within and outside of an organizati...

Description

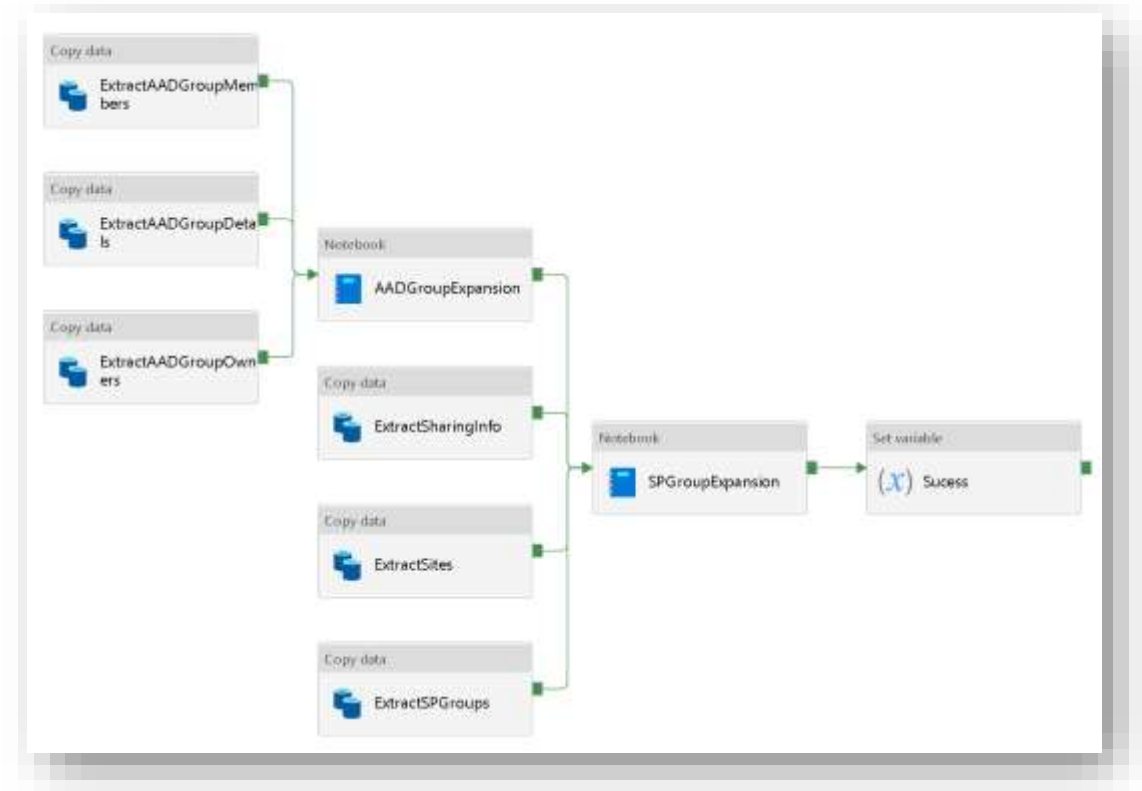
Use this template to accelerate sharing security scenarios by identifying information sharing within and outside of an organization. This template extracts Microsoft 365 (Office) SharePoint data via Microsoft Graph Data Connect and aggregates with Azure Active Directory groups to produce analytics-ready data for analysis.

We would love to hear your thoughts on this template. Please send us your ideas and feedback at <https://aka.ms/synapse-m365-sharepoint-feedback>.

[View documentation](#)

Tags

- MGDC
- Azure Synapse Analytics
- OneDrive
- SharePoint
- Security
- AAD
- Sharing
- Sites
- SPGroups
- Documents
- Syntex
- M365
- Office365
- Graph
- Microsoft
- Office

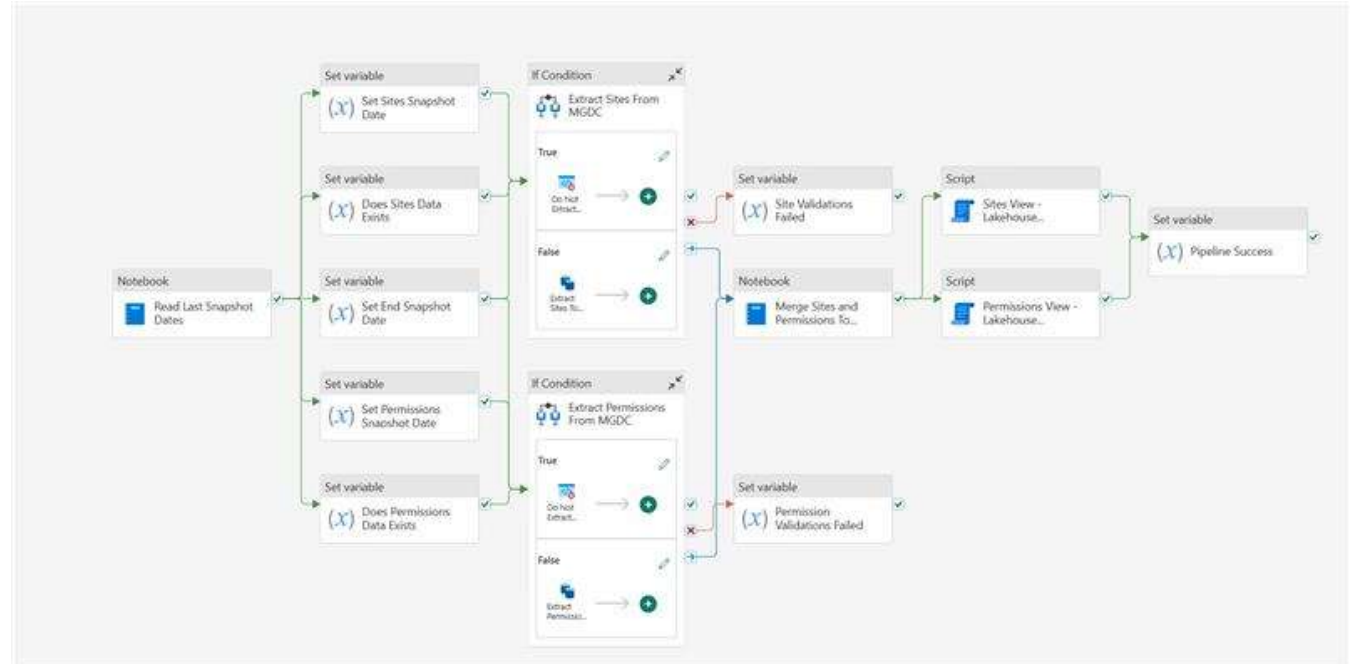


Information Oversharing Template Setup



Oversharing Template v2 - Microsoft Fabric

- Latest MGDC for SharePoint template
- Main improvements:
 - Uses Microsoft Fabric for improved scalability and simpler SQL queries.
 - Uses the new UserCount columns to find the number of users in each group.
 - No longer needs SharePoint Groups and Microsoft Entra ID Groups datasets.
 - Implements delta datasets for Sites and Permissions.
 - Flattens the permission dataset, making it easier to query permission details.
- This new template reduces the required resources for tenants interested in the Oversharing scenario.

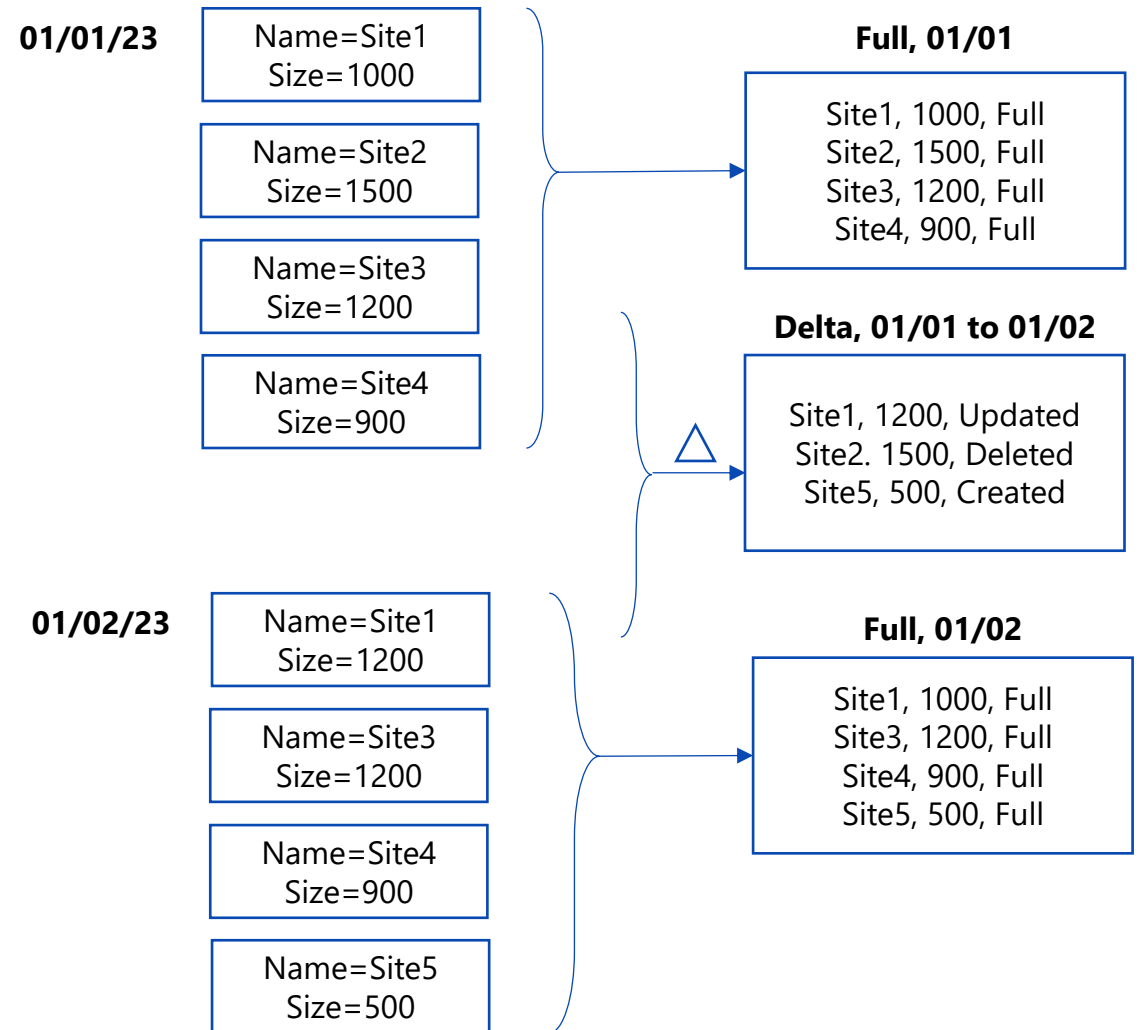


<https://aka.ms/fabricoversharingtemplatesteps>



Delta Datasets

- What is a Delta Dataset?
 - Partial dataset with only objects that changed
 - Objects created, modified, or deleted in the period
 - Potential to reduce the daily transfers dramatically
 - Amount of reduction depends on usage
 - Must process delta to update the last full copy
 - Offered for SharePoint datasets (Sites, Groups and Permissions)
- Delta Advantages
 - Fewer objects transferred for dataset updates
 - More optimal use of resources
 - Smaller Azure bill
- Delta Disadvantages
 - More complexity to keep a State Dataset updated
 - Increased processing to update a State Dataset



Consumption-based Model

- Azure subscription required (where data is stored, where pipelines run)
- Consumption charged per object delivered (shows as part of the Azure bill)
- Azure storage costs depend on frequency, data size and retention policy
- Delta pulls reduce the cost of large state datasets
- Aligned with Microsoft Graph Data Connect
 - Billed via Azure at \$0.75 per 1,000 objects
 - This is the list price. Azure discounts may apply
 - See [Pricing – MGDC | Microsoft Azure](#)
- Important Note: The Files dataset is billed at a special rate of \$0.75 per 50,000 objects. That is the only MGDC dataset billed differently.

Estimated cost for pulling 50K objects

| Azure | Full pull 1 day | Full pull 30 days | Delta* 30 days |
|--------------------|-----------------|-------------------|----------------|
| SharePoint Data | \$37.50 | \$1,125.00 | \$112.50 |
| Azure Data Factory | \$0.75 | \$22.50 | \$22.50 |
| Azure Storage | \$0.00 | \$0.03 | \$0.03 |
| Total | \$38.75 | \$1,147.53 | \$135.03 |

* = assumes 10% of the objects change every day and that a previous full pull is available



Consumption-based Model

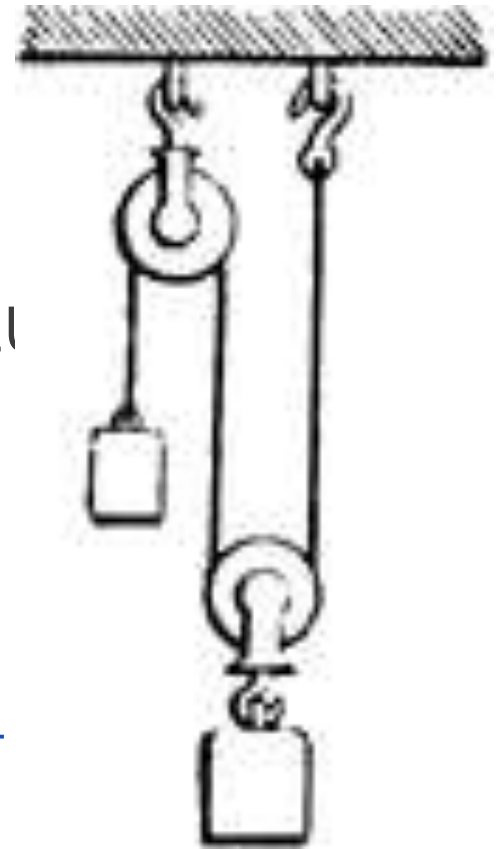
| Collab Level* | Sites | Groups | Permissions | Files | Sites Cost/Pull | Groups Cost/Pull | Permissions Cost/Pull | Files Cost/Pull | Security** Cost/Pull | Capacity** Cost/Pull | Security*** Cost/Delta | Capacity*** Cost/Delta |
|---------------|-----------|------------|-------------|---------------|-----------------|------------------|-----------------------|-----------------|----------------------|----------------------|------------------------|------------------------|
| Average | 5,000 | 155,000 | 305,000 | 14,370,000 | \$ 4 | \$ 116 | \$ 229 | \$ 216 | \$ 349 | \$ 219 | \$ 17 | \$ 11 |
| Average | 10,000 | 310,000 | 610,000 | 28,740,000 | \$ 8 | \$ 233 | \$ 458 | \$ 431 | \$ 698 | \$ 439 | \$ 35 | \$ 22 |
| Average | 20,000 | 620,000 | 1,220,000 | 57,480,000 | \$ 15 | \$ 465 | \$ 915 | \$ 862 | \$ 1,395 | \$ 877 | \$ 70 | \$ 44 |
| Average | 50,000 | 1,550,000 | 3,050,000 | 143,700,000 | \$ 38 | \$ 1,163 | \$ 2,288 | \$ 2,156 | \$ 3,488 | \$ 2,193 | \$ 174 | \$ 110 |
| Average | 100,000 | 3,100,000 | 6,100,000 | 287,400,000 | \$ 75 | \$ 2,325 | \$ 4,575 | \$ 4,311 | \$ 6,975 | \$ 4,386 | \$ 349 | \$ 219 |
| Average | 500,000 | 15,500,000 | 30,500,000 | 1,437,000,000 | \$ 375 | \$ 11,625 | \$ 22,875 | \$ 21,555 | \$ 34,875 | \$ 21,930 | \$ 1,744 | \$ 1,097 |
| Average | 1,000,000 | 31,000,000 | 61,000,000 | 2,874,000,000 | \$ 750 | \$ 23,250 | \$ 45,750 | \$ 43,110 | \$ 69,750 | \$ 43,860 | \$ 3,488 | \$ 2,193 |
| Higher | 5,000 | 310,000 | 610,000 | 28,740,000 | \$ 4 | \$ 233 | \$ 458 | \$ 431 | \$ 694 | \$ 435 | \$ 69 | \$ 43 |
| Higher | 10,000 | 620,000 | 1,220,000 | 57,480,000 | \$ 8 | \$ 465 | \$ 915 | \$ 862 | \$ 1,388 | \$ 870 | \$ 139 | \$ 87 |
| Higher | 20,000 | 1,240,000 | 2,440,000 | 114,960,000 | \$ 15 | \$ 930 | \$ 1,830 | \$ 1,724 | \$ 2,775 | \$ 1,739 | \$ 278 | \$ 174 |
| Higher | 50,000 | 3,100,000 | 6,100,000 | 287,400,000 | \$ 38 | \$ 2,325 | \$ 4,575 | \$ 4,311 | \$ 6,938 | \$ 4,349 | \$ 694 | \$ 435 |
| Higher | 100,000 | 6,200,000 | 12,200,000 | 574,800,000 | \$ 75 | \$ 4,650 | \$ 9,150 | \$ 8,622 | \$ 13,875 | \$ 8,697 | \$ 1,388 | \$ 870 |
| Higher | 500,000 | 31,000,000 | 61,000,000 | 2,874,000,000 | \$ 375 | \$ 23,250 | \$ 45,750 | \$ 43,110 | \$ 69,375 | \$ 43,485 | \$ 6,938 | \$ 4,349 |
| Higher | 1,000,000 | 62,000,000 | 122,000,000 | 5,748,000,000 | \$ 750 | \$ 46,500 | \$ 91,500 | \$ 86,220 | \$ 138,750 | \$ 86,970 | \$ 13,875 | \$ 8,697 |

Source: [MGDC for SharePoint FAQ: How can I estimate my Azure bill?](#)



Public Preview

- 3 Scenarios: Security, Capacity and Sync Health
 - Available in 15 regions: APAC, ARE, AUS, BRA, CAN, CHE, DEU, EIM, FRA, GBR, IND, JPN, KOR, NAM, NOR
 - Column filtering, row filtering and sampling
 - Allow 4 days after signup for data to become available
 - Details and Setup instructions at <https://aka.ms/SharePointData>
 - Billed via Azure at \$0.75 per 1,000 objects*
 - This is the list price. Your Azure discounts apply
- * Exception: the Files dataset is billed at \$0.75 per 50,000 objects



Thank You!



Steps to Setup and Use the Microsoft Graph Data Connect with the Azure Synapse Copy Data Tool



Org Settings

The screenshot shows the Microsoft 365 admin center interface. The top navigation bar includes the Microsoft 365 admin center logo, a search bar, and utility icons for mail, documents, notifications, and a user profile (MC). The left sidebar contains a menu with categories like Home, Users, Teams & groups, Roles, Resources, Billing, Support, and Settings. Under the Settings category, 'Org settings' is selected and highlighted with a blue bar. The main content area displays a list of settings under the 'Services' tab. The list includes items like Microsoft Edge product messaging, Microsoft Edge site lists, Microsoft Forms, Microsoft Graph Data Connect, Microsoft Loop Preview, Microsoft Planner, Microsoft Search in Bing homepage, and Microsoft Teams. The 'Microsoft Graph Data Connect' item is highlighted with a grey background and a blue arrow pointing to it from the right.

Microsoft 365 admin center

Search

Services Security & privacy Organization profile Search all settings

41 items 1 selected X

| Name ↑ | Description |
|--|--|
| Microsoft Edge product messaging for users | Accelerate your Microsoft Edge adoption goals by introducing |
| Microsoft Edge site lists | Set websites to open in Internet Explorer mode in Microsoft |
| Microsoft Forms | Manage external sharing and record the names of people in |
| Microsoft Graph Data Connect | Manage and update your Microsoft Graph Data Connect set |
| Microsoft Loop Preview | Control access to Microsoft Loop workspaces. |
| Microsoft Planner | Choose whether your users can publish plans and assigned t |
| Microsoft Search in Bing homepage | Customize Bing.com page for your signed-in users. |
| Microsoft Teams | Manage and update your Microsoft Teams settings. |



Enable MGDC, Select Approvers Group, Enable SharePoint and OneDrive Datasets

Microsoft 365 admin center

Search

Microsoft Graph Data Connect

Microsoft Graph Data Connect enables bulk transfer of data to Azure for use by applications you select for use in your organization. Your applications are governed by your organization's policies and the applicable separate agreement(s) you enter into with the third party who provides the applications (if any). Microsoft is not responsible for use, inability to use, or performance of products or services not provided by Microsoft.

To get started with Microsoft Graph Data Connect, please see the [Azure Data Factory or Azure Synapse Analytics documentation of the Office 365 connector](#). You can approve or deny applications using [Microsoft Graph Data Connect applications portal](#).

[Azure Data Factory or Azure Synapse Analytics documentation of the Office 365 connector](#)

[Microsoft Graph Data Connect applications portal \(app consent\)](#)

Turn Microsoft Graph Data Connect on or off for your entire organization.

Enable dataset access from these services. [Learn More](#)

Viva Insights

SharePoint Online and OneDrive

Note: It can take up to 96 hours for SharePoint Online and OneDrive datasets to be accessible after you enable this option.

Note: Enabling Microsoft Graph Data Connect is optional. If enabled, Microsoft Graph Data Connect will not honor M365 Customer Key encryption policies when transferring your data to your organization's Microsoft Azure data store, and data that is transferred to Microsoft Azure will honor storage policies setup in Microsoft Azure.

[Learn More](#)

Enable MGDC

Enable SharePoint and OneDrive datasets



Register an Application

The image shows a sequence of three screenshots from the Microsoft Azure portal, illustrating the steps to register an application and manage its credentials.

Top Screenshot: Archimedes | App registrations
This is the main page for app registrations in the Azure Active Directory for 'Archimedes'. The 'New registration' button is highlighted with a white arrow.

Middle Screenshot: Register an application
This is the 'Register an application' form. The 'Name' field is visible, and the 'Supported account types' section is expanded. The 'Accounts in this organizational directory only (Admins and Users)' option is selected. A white arrow points to the 'Supported account types' section.

Bottom Screenshot: JoseTestApp | Certificates & secrets
This is the 'Certificates & secrets' page for the 'JoseTestApp'. The 'Client secrets (1)' tab is active, showing a table of secrets. A white arrow points to the 'Certificates & secrets' menu item in the left-hand navigation pane.

Client secrets table:

| Description | Expires | Value | Secret ID |
|-------------|-----------|------------------------|---------------------------|
| AppSecret | 3/22/2023 | H7X8Q~slwcyOeW4SqXz... | 81601454-6a4e-4bb5-b20... |

Create Storage Account and Assign Role

The screenshot shows the 'Create storage account' page in the Microsoft Azure portal. The breadcrumb navigation is 'Home > Storage accounts >'. The page title is 'Create storage account'. There are tabs for 'Basics', 'Networking', 'Data protection', 'Advanced', 'Tags', and 'Review + create'. The 'Basics' tab is selected. The page contains the following sections:

- Project details:** Subscription is 'Azure subscription 1' and Resource group is '(New) RG_DataConnect'. There is a 'Create new' link for the resource group.
- Instance details:** Storage account name is 'm365x915006storage', Location is '(US) East US', Performance is 'Standard', Account kind is 'StorageV2 (general purpose v2)', and Replication is 'Locally-redundant storage (LRS)'.

At the bottom, there are three buttons: 'Review + create', '< Previous', and 'Next : Networking >'. A blue arrow points from the 'Storage Blob Data Contributor' role in the second screenshot to the 'Next : Networking >' button in this screenshot.

The screenshot shows the 'Access Control (IAM)' page for the storage account 'm365x915006storage'. The breadcrumb navigation is 'Home > m365x915006storage'. The page title is 'm365x915006storage | Access Control (IAM)'. There are tabs for 'Check access', 'Role assignments', 'Roles', 'Deny assignments', and 'Classic administrators'. The 'Role assignments' tab is selected. The page shows the following information:

- Number of role assignments for this subscription: 2 / 2000
- Search by name or email: [Search box]
- Filters: Type: All, Role: All, Scope: All scopes, Group by: Role
- 2 items (1 Users, 1 Service Principals)

| <input type="checkbox"/> | Name | Type | Role | Scope |
|--------------------------|---|------|-------------------------------|--------------------------|
| <input type="checkbox"/> | Owner | | | |
| <input type="checkbox"/> | MOD Administrator admin@m365x915006.onmicrosoft.com | User | Owner | Subscription (Inherited) |
| <input type="checkbox"/> | Storage Blob Data Contributor | | | |
| <input type="checkbox"/> | Microsoft Graph Data Connect Data Transfer | App | Storage Blob Data Contributor | This resource |

A blue arrow points to the 'Storage Blob Data Contributor' role assignment for the 'Microsoft Graph Data Connect Data Transfer' app.



Add your Graph Data Connect application in Azure

Microsoft Azure

Search resources, services, and docs (G+/)

Home >

Graph Data Connect

+ Add Refresh Delete

Search Application State == All

| Name ↑ | Application ID | Registered On | Developer |
|--------|----------------|---------------|-----------|
|--------|----------------|---------------|-----------|

Get started with Microsoft Graph Data Connect to access rich data at scale

The following policies and statements apply while the service is in Private Preview:

- Microsoft Privacy Statement
- Microsoft Graph Data Connect
- Microsoft Preview policy

Add a new application

Microsoft Azure

Search resources, services, and docs (G+/)

Home > Graph Data Connect >

Data Connect

Registration Info Review + create

Add an application for Microsoft Graph Data Connect

Add a Microsoft Graph Data Connect application to access rich data from Microsoft 365 datasets at scale. An application needs to be associated to a Subscription, a Region, an Active Directory app, and a Storage Account. Additional characteristics for Publish Type and Datasets enable traceability and governance. The following policies and statements apply while the service is in Private Preview: [Microsoft Privacy Statement](#), [Microsoft Graph Data Connect](#), [Microsoft Preview policy](#)

Project Details

Select the subscription to manage deployed resources and costs.

Subscription *

Resource Group *

Destination Type

Storage Account *

Storage Account URI *

Instance Details

Select a publishing type, storage account, and the datasets to associate to the application.

Application ID *

Description *

Publish Type Multi-Tenant Single-Tenant

Review + Create < Previous Next > Datasets >

Home > Graph Data Connect >

Data Connect

Registration Info **Datasets** Re

Datasets are table/columns pairs that [Learn more about datasets](#) and

Dataset

- BasicDataSet_v0.Message_v0
- BasicDataSet_v0.Message_v1
- BasicDataSet_v0.OutlookGroupConversations_v0
- BasicDataSet_v0.PlannerTasks_v0
- BasicDataSet_v0.Rooms_v0
- BasicDataSet_v0.SentItem_v0
- BasicDataSet_v0.SentItem_v1
- BasicDataSet_v0.SharePointGroups_v1
- BasicDataSet_v0.SharePointPermissions_v1
- BasicDataSet_v0.SharePointSites_v1
- BasicDataSet_v0.TeamChat_v1
- BasicDataSet_v0.TeamChat_v2

Approve the application in the Admin Center

Microsoft 365 admin center

Home > Org settings

Org settings

Services Security & privacy Organization profile

Search all settings

- Home
- Users
- Teams & groups
- Roles
- Resources
- Billing
- Support
- Settings
- Domains
- Search & intelligence
- Org settings
- Integrated apps
- Partner relationships
- Microsoft Edge
- Setup

Microsoft Graph Data Connect applications

Home > Org settings > Microsoft Graph Data Connect applications

Microsoft Graph Data Connect applications

Apps

This is the list of Microsoft Graph Data Connect (MGDC) applications which require Admin approval to run. MGDC is a product which enables mass data egress for data analytics and app development. Learn more [here](#).

+ Find new multi-tenant app

| App name | App type | Last consented d... | Description |
|----------|---------------|---------------------|-------------|
| MGDC_App | Single-tenant | | MGDC App |

Microsoft 365 admin center

MGDC_App

- Overview
- BasicDataSet_v0.SharePointSib
- Review**

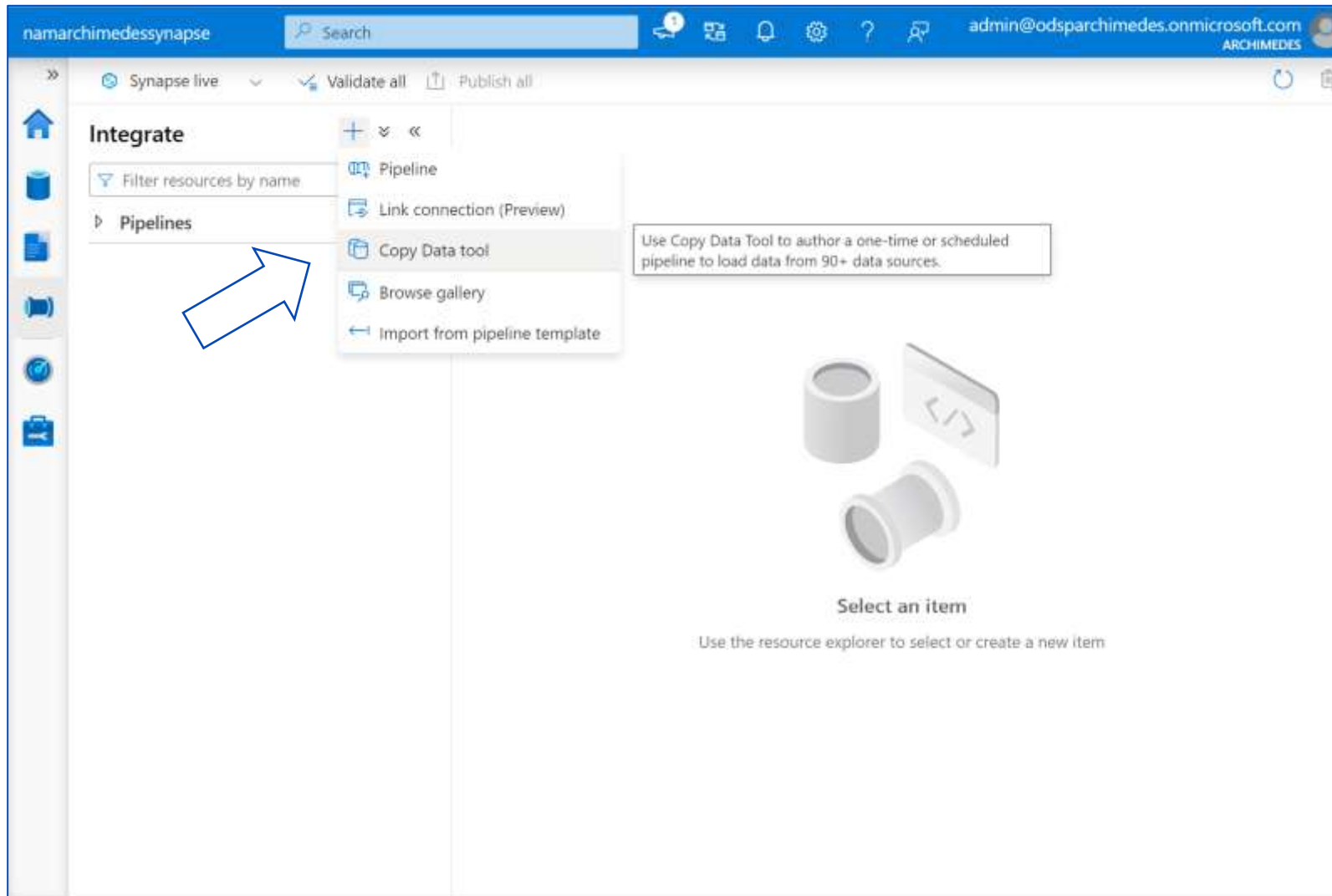
Review

Data destination
https://[redacted].dfs.core.windows.net/

App publisher
3ac-[redacted]-cb3

Back Approve Decline Cancel

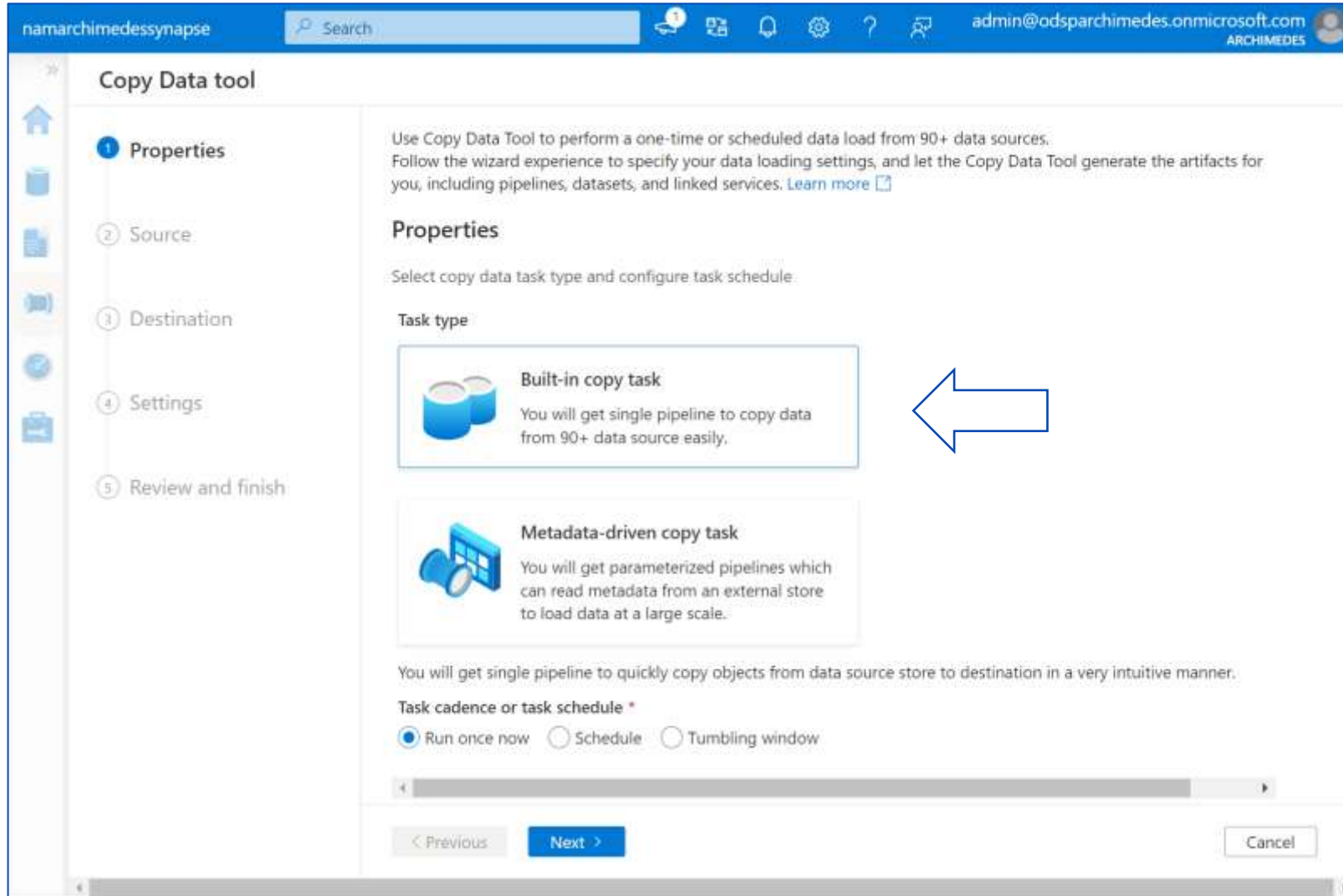
In the Synapse Analytics Workspace, start the Copy Data Tool



In the Synapse Analytics Workspace, open “Synapse Studio”, select “Integrate” to see the pipelines, select “+” to create a new pipeline and use the option for “Copy Data tool”



Select the “Built-in copy task”



The screenshot shows the 'Copy Data tool' configuration wizard in the Azure portal. The interface is in a blue header with the user 'admin@odsparchimedes.onmicrosoft.com' and the 'ARCHIMEDES' logo. The left sidebar shows a progress indicator with five steps: 1. Properties (selected), 2. Source, 3. Destination, 4. Settings, and 5. Review and finish. The main content area is titled 'Copy Data tool' and contains the following text:

Use Copy Data Tool to perform a one-time or scheduled data load from 90+ data sources. Follow the wizard experience to specify your data loading settings, and let the Copy Data Tool generate the artifacts for you, including pipelines, datasets, and linked services. [Learn more](#)

Properties

Select copy data task type and configure task schedule.

Task type

- Built-in copy task**
You will get single pipeline to copy data from 90+ data source easily.
- Metadata-driven copy task**
You will get parameterized pipelines which can read metadata from an external store to load data at a large scale.

You will get single pipeline to quickly copy objects from data source store to destination in a very intuitive manner.

Task cadence or task schedule *

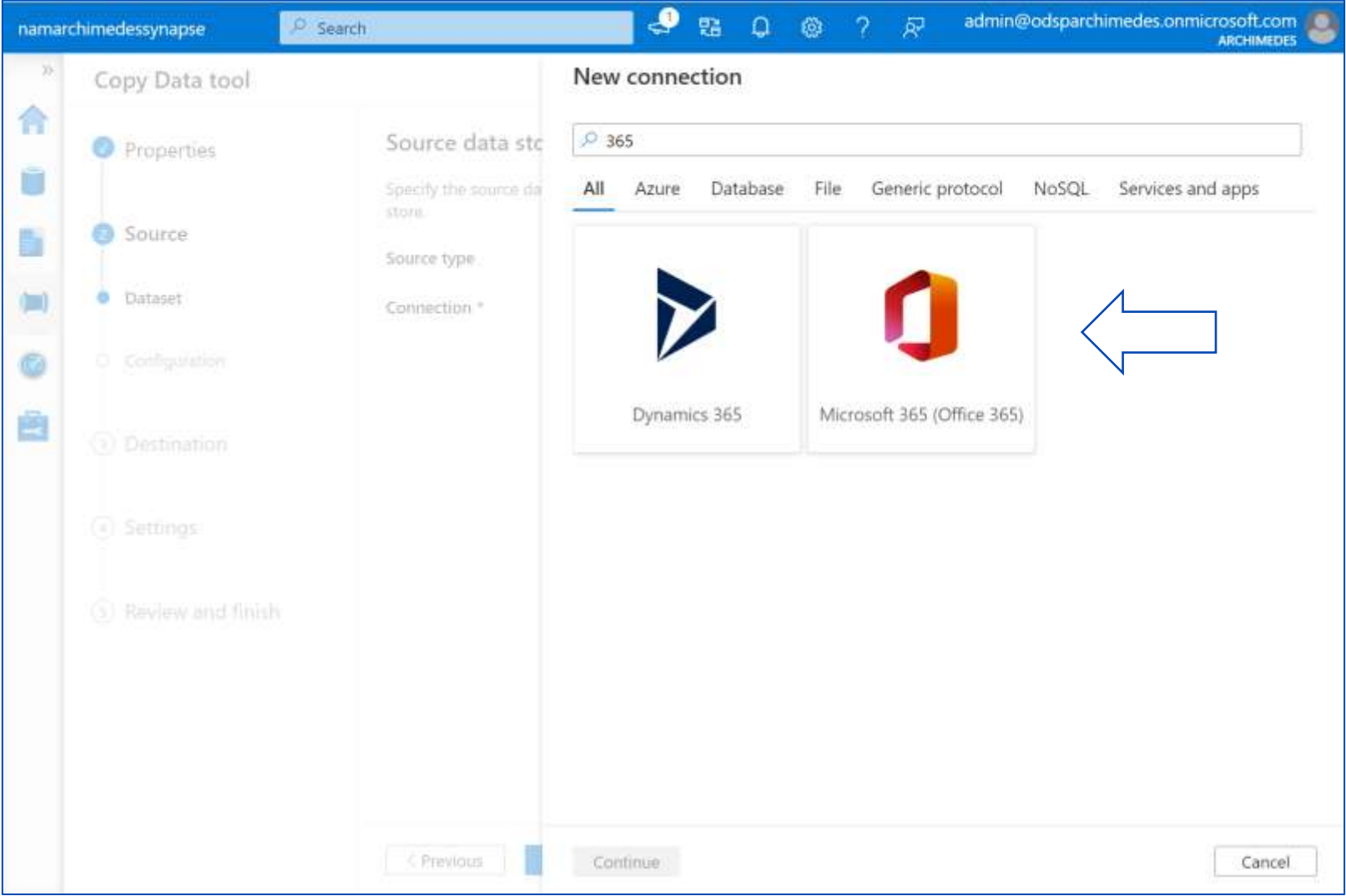
Run once now Schedule Tumbling window

Navigation buttons at the bottom: < Previous, Next >, and Cancel.

A blue arrow points to the 'Built-in copy task' option.



Create a new connection to Microsoft 365



Enter a connection name, application id and application credentials

The screenshot shows the 'New connection' configuration page in the Copy Data tool. The page is titled 'New connection' and is for a 'Microsoft 365 (Office 365)' connection. The 'Name' field is filled with 'Microsoft3651'. The 'Description' field is empty. The 'Connect via integration runtime' dropdown is set to 'AutoResolveIntegrationRuntime'. The 'Microsoft 365 Tenant' dropdown is set to 'Archimedes (3a3630bf-4d67-4c21-842a-abcea48840d5)'. The 'Service principal ID' field is empty. The 'Service principal key' section is selected, and the 'Service principal key' field is empty. The 'Annotations' field is empty. The 'Create' button is highlighted.

Copy Data tool

Properties

Source

Dataset

Configuration

Destination

Settings

Review and finish

Source data store

Specify the source data store.

Source type

Connection *

New connection

Microsoft 365 (Office 365) Learn more

Choose a name for your linked service. This name cannot be updated later.

Name *

Microsoft3651

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Microsoft 365 Tenant *

Archimedes (3a3630bf-4d67-4c21-842a-abcea48840d5)

Service principal ID *

Service principal key Azure Key Vault

Service principal key *

Annotations

Previous Create Back Test connection Cancel



Select an existing connection, then select one of the datasets

Copy Data tool

Properties

2 Source

Dataset

Configuration

3 Destination

5 Review and finish

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type: All

Connection: MGDC_Source Edit + New connection

Integration runtime: AutoResolveIntegrationRuntime Edit

Filter by name... Refresh Showing 36 out of 36 items (1 selected)

- BasicDataSet_v0.SentItem_v1
- BasicDataSet_v0.SharePointGroups_v1
- BasicDataSet_v0.SharePointPermissions_v1
- BasicDataSet_v0.SharePointSites_v1
- BasicDataSet_v0.TeamChat_v1

< Previous **Next >** Cancel



Select the SnapshotDate as a filter, choosing a start date and end date

Copy Data tool

- Properties
- Source
- Dataset
- Configuration
- Destination
- Settings
- Review and finish

Apply filter

Choose date time column that will be used to filter data or use a custom query. ⓘ

Filter by name...
BasicDataSet_v0.SharePointSites_v1

Scope: All users in the Microsoft 365 tenant

Scope filter

| Column name * | Start time (UTC) * | End time (UTC) * |
|---------------|--------------------|------------------|
| SnapshotDate | 10/7/22 00:00:00 | 10/7/22 00:00:00 |

< Previous Next > Cancel



Enter a connection name, authentication type (Service Principal), subscription and storage account

The screenshot shows the 'New connection' configuration window in the Copy Data tool. The window is titled 'New connection' and is for 'Azure Data Lake Storage Gen2'. The user is 'admin@odsparchimedes.onmicrosoft.com'.

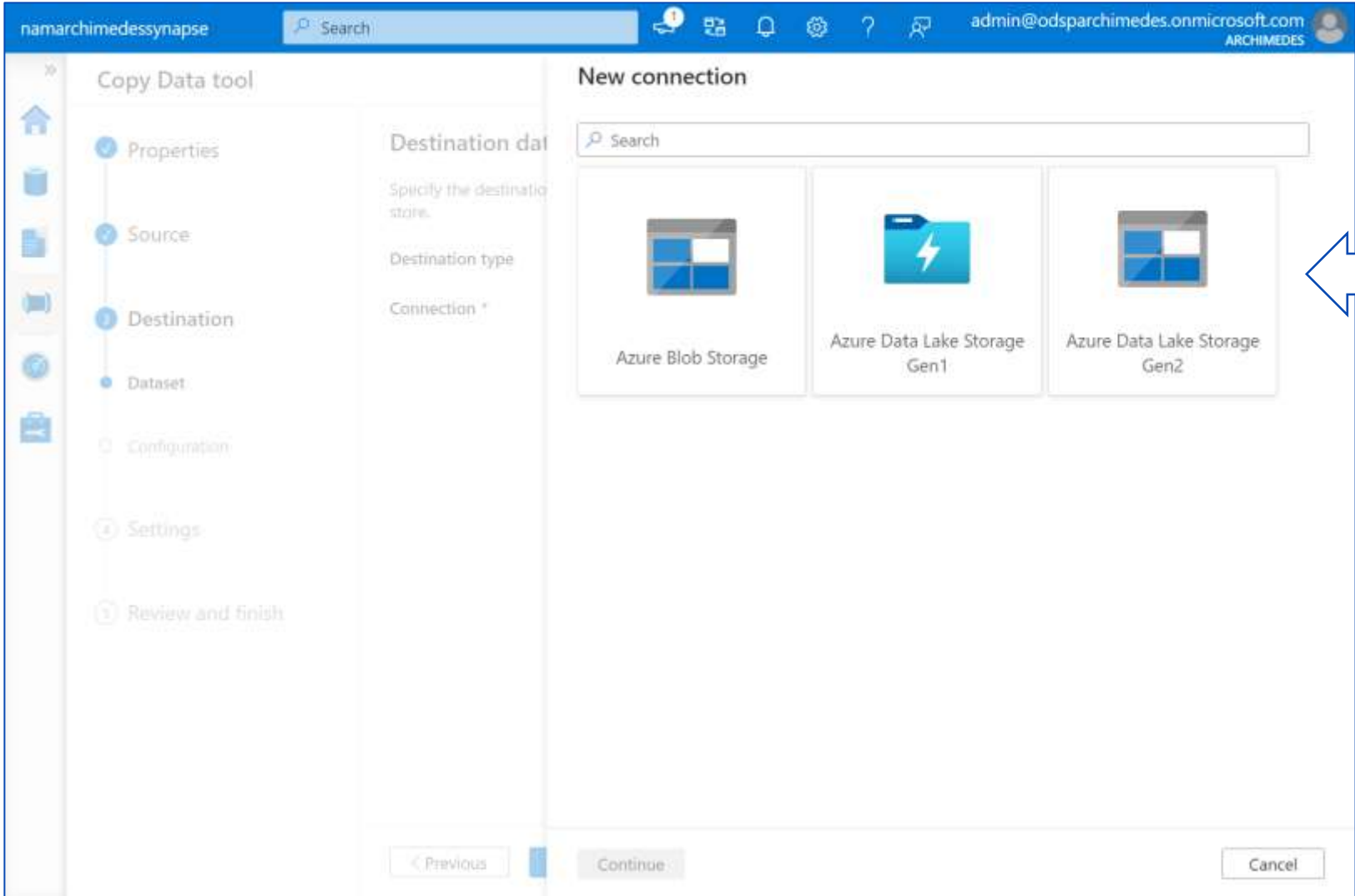
The configuration steps are:

- Name ***: AzureDataLakeStorage1
- Description**: (Empty text area)
- Connect via integration runtime ***: AutoResolveIntegrationRuntime
- Authentication type**: Service Principal
- Account selection method**: From Azure subscription (selected)
- Azure subscription**: odsparchimedes (f48ff060-e455-4ca8-8243-e644ce623d6b)
- Storage account name ***: cs71003200184be2293

Buttons at the bottom: Previous, Create, Back, Test connection, Cancel.



Create a new data store connection, select ADLS Gen2



Select an existing data store connection and enter a path

The screenshot shows the 'Copy Data tool' configuration interface. The left sidebar contains a navigation pane with the following steps: Properties, Source, Destination (selected), Dataset, Configuration, Settings, and Review and finish. The main content area is titled 'Destination data store' and includes the following fields:

- Destination type:** A dropdown menu set to 'All'.
- Connection *:** A dropdown menu set to 'LS_ADLSGen2', with 'Edit' and '+ New connection' options.
- Integration runtime *:** A dropdown menu set to 'AutoResolveIntegrationRuntime', with an 'Edit' option.
- Folder path:** A text input field containing 'newcontainer/singlesite', with a 'Browse' button below it.
- File name:** An empty text input field.
- Max concurrent connections [ⓘ]:** An empty text input field.
- Block size (MB) [ⓘ]:** An empty text input field.

At the bottom of the configuration area, there are three buttons: '< Previous', 'Next >', and 'Cancel'.



Enter a name for the pipeline

The screenshot displays the 'Copy Data tool' configuration page in Microsoft Synapse Studio. The left sidebar shows the navigation menu with 'Review and finish' selected. The main area is divided into 'Summary', 'Properties', 'Source', and 'Destination' sections.

Summary: You are running pipeline to copy data from Microsoft 365 (Office 365) to Azure Data Lake Storage Gen2.

Diagram: Microsoft 365 (Office 365) → Azure Data Lake Storage Gen2

Properties:

- Task name:
- Task description:

Source:

- Connection name: LS_Office365
- Dataset name: SourceDataset_ya4
- Number of tables: SharePointSitesDataset_v0_Preview

Destination:

Navigation buttons: < Previous, Next >, Cancel



Let the pipeline deploy and run. Monitor it.

The screenshot shows the 'Copy Data tool' interface in a web browser. The top navigation bar includes the user name 'namarchimedessynapse', a search bar, and the email 'admin@odsparchimedes.onmicrosoft.com'. The left sidebar contains a vertical list of steps: Properties, Source, Destination, Settings, Review and finish (highlighted with a blue circle and the number 5), Review, and Deployment. The main content area features a diagram with the Microsoft 365 (Office 365) logo on the left and the Azure Data Lake Storage logo on the right, connected by a right-pointing arrow. Below the diagram, the text 'Deployment complete' is displayed. A table lists the deployment steps and their status:

| Deployment step | Status |
|-------------------------------------|-------------|
| Validating copy runtime environment | ✔ Succeeded |
| > Creating datasets | ✔ Succeeded |
| > Creating pipelines | ✔ Succeeded |
| > Running pipelines | ✔ Succeeded |

Below the table, a message states: 'Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.' At the bottom of the interface, there are three buttons: 'Finish' (highlighted in blue), 'Edit pipeline', and 'Monitor'.



Monitor the pipeline

The screenshot displays the Microsoft Azure Data Factory interface for monitoring pipeline runs. The main view shows the 'Pipeline runs' section for the 'Sites_Pipeline'. The interface includes a navigation pane on the left with categories like Analytics pools, Activities, and Integration. The main content area shows a table of pipeline runs with columns for Pipeline name, Run start, Run end, Duration, Triggered by, and Status. A detailed view of the 'Copy_j1y' activity is also shown, displaying its status as 'ExtractingData'.

Pipeline runs

Triggered Debug Rerun Cancel options Refresh Edit columns List Gantt

Filter by run ID or name Local time: Last 24 hours Pipeline name: Sites_Pipeline

Status: Succeeded Runs: Latest runs Triggered by: All Add filter

Showing 1 - 1 items Last refreshed 0 minutes ago

| Pipeline name | Run start | Run end | Duration | Triggered by | Status |
|----------------|--------------------------|--------------------------|----------|----------------|-----------|
| Sites_Pipeline | Oct 13, 2022, 6:36:58 pm | Oct 13, 2022, 7:08:15 pm | 00:31:17 | Manual trigger | Succeeded |

Activity runs

Pipeline run ID eb8d5dbf-276f-4525-a7a9-75feafa8ab4c

All status

Showing 1 - 1 items

| Activity name | Activity type | Run start | Duration | Status | Error |
|---------------|---------------|-----------------------|----------|----------------|-------|
| Copy_j1y | Copy data | Oct 13, 2022, 9:38:18 | 01:46:10 | ExtractingData | |

