

Export to Data Lake Preview

Contents

Announcing Export to data lake (Preview)	1
Preview restrictions apply.....	1
Overview	1
Pre-requisites for using the Export to Data Lake service.....	2
Key features	2
How to export Common Data Service entities to an Azure data lake in your subscription:	3
1. Select the storage account in your Azure subscription	3
2. Add entities.....	4
Support for initial and incremental writes for data and metadata	5
Support for replicating create, update and delete operations.....	Error! Bookmark not defined.
Continuous snapshot updates for large analytics scenarios.....	Error! Bookmark not defined.

Announcing Export to data lake (Preview)

We are super excited to announce the Export to data lake preview to our Common Data Service customers. The Export to data lake service enables continuous replication of Common Data Service entity data to Azure data lake which can then be used to run analytics such as Power BI reporting, ML, data warehousing or other downstream integration purposes. It simplifies the technical and administrative complexity of operationalizing entities for analytics and managing schema and data. Within a matter of minutes, customers will be able to link their Common Data Service environment to a data lake in their Azure subscription, select standard or customer entities and export it to data lake. Any data or metadata changes (initial and incremental) in the Common Data Service is automatically pushed to the lake without any additional actions.

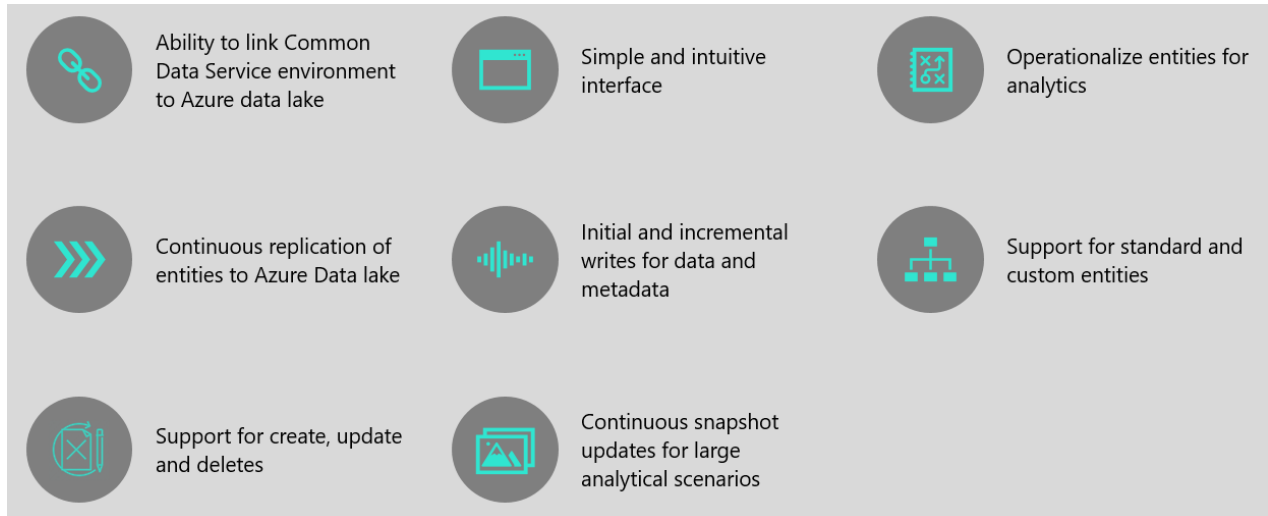
Preview restrictions apply

- Preview features aren't meant for production use, may have restricted functionality and are subject to a separate supplemental term of use
- Preview features are available before an official release so that customers can get early access and provide feedback.
- We expect changes to this feature, so you shouldn't use it in production. Use it only in test, demo or development environments.
- Microsoft doesn't provide support for this preview feature

Overview

Our vision is to empower our customers to gain comprehensive insights and drive business actions based on their data in the Common Data Service (CDS). To enable this, we are building a new service

called Export to data lake which is a pipeline to continuously export data from the Common Data Service to Azure data lake; designed for enterprise big data analytics, cost-effective, scalable, has high availability/disaster recover capabilities and enables best in class analytics performance. Data is stored in the Common Data Model (CDM) format which provides semantic consistency across apps and deployments. The standardized metadata and self-describing data in an Azure data lake facilitates metadata discovery and interoperability between data producers and consumers such as Power BI, Azure Data Factory, Azure Databricks, and Azure Machine Learning service.



Pre-requisites for using the Export to Data Lake service

Before you can export Common Data Service data to a data lake, you must create and configure an Azure data lake Gen 2 storage account:

- Follow the steps in the [Create an Azure Data Lake Storage Gen2 storage account](#) article
- The storage account must be created in the same Azure AD tenant as your PowerApps tenant.
- Set your storage as Storagev2 (general purpose v2)
- The storage account must have the Hierarchical Name Space feature enabled
- You must be granted an Owner role on the storage account

Note:

- It is recommended that the storage account is created in the same region as the PowerApps environment you plan to use it in.
- It is recommended to set replication setting to Read-access geo-redundant storage (RA-GRS)

Key features

- Ability to link/unlink the Common Data Service environment to a data lake in customer's Azure subscription
- Continuous replication of entities to Azure data lake
- Support for initial and incremental writes for data and metadata
- Simple and intuitive interface
- Support for replicating standard and custom entities
- Support for replicating create, update and delete operations
- Continuous snapshot updates for large analytics scenarios

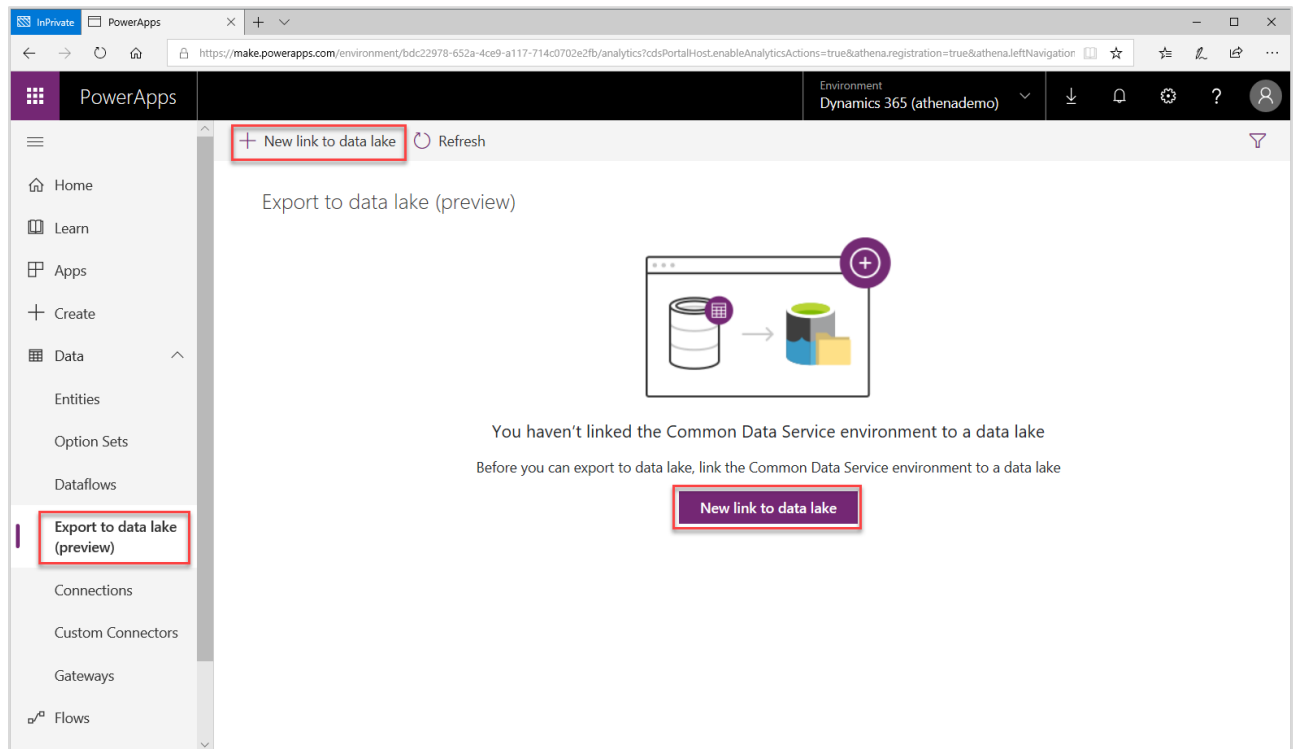
How to export Common Data Service entities to an Azure data lake in your subscription:

To export your entity data to a data lake, you need to first link your Common Data Service environment to a data lake in your Azure subscription. A simple and intuitive user interface walks you through the 'New link to data lake' wizard wherein you will be asked to select your Azure storage account and then add entities that you want to replicate to the lake.

1. Select the storage account in your Azure subscription

[Click this link](#) to go to the PowerApps portal (Note: This link includes a feature flag as the service is not public yet)

From the PowerApps maker portal, select **Export to data lake** service in the left-hand pane and launch the **New link to data lake** wizard



At the **Select Storage Account** step, pick your Azure subscription and resource group and then select the storage account that you want to link to the Common Data Service environment

PowerApps

Environment
Dynamics 365 (athenademo)

Home

Learn

Apps

Create

Data

Entities

Option Sets

Dataflows

Export to data lake (preview)

Connections

Custom Connectors

Gateways

Flows

New link to data lake

Select Storage Account
athenademo

Add Entities

Select Storage Account

Select the storage account that you want link to the Common Data Service environment.

As part of linking the Common Data Service environment to a data lake, you are granting the Export to data lake service additional roles to your storage account.

Subscription *

Athena Demo

Resource group *

athena

Storage account *

athenademo

Back

Next

Cancel

Note:

Please note that as part of linking the Common Data Service environment to a data lake, you are granting the Export to data lake service access to your storage account.

Additionally, you are also granting the Power Platform Dataflows service access to your storage account. For more information, please refer to the [Dataflows documentation](#)

Next step is to select the entities you want to export Azure data lake.

2. Add entities

PowerApps

Environment
Dynamics 365 (athenademo)

Home

Learn

Apps

Create

Data

Entities

Option Sets

Dataflows

Export to data lake (preview)

Connections

Custom Connectors

Gateways

Flows

New link to data lake

Select Storage Account
athenademo

Add Entities
143 of 143 selected

Add Entities

Select the entities that you want to enable for analytics.

Filter by keyword

Entity	Name
Account	account
Action Card	actioncard
Activity Party	activityparty
Actual	msdyn_actual
Agreement Business Process	msdyn_bpf_baa0a411a239410cb8bde8b5fdd88e3
Analysis Component	msdyn_analysiscomponent
Analysis Job	msdyn_analysisjob

Back

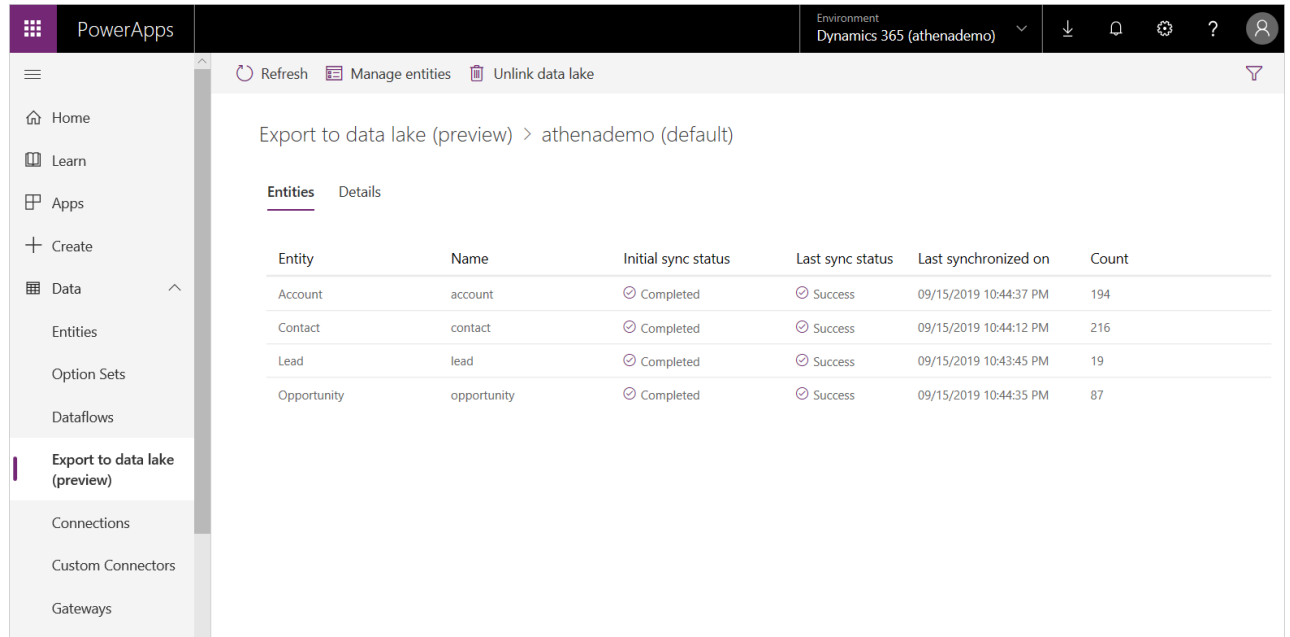
Save

Cancel

At the Add entities step, select the Common Data Service entities whose data you want to push to the lake.

After you hit 'Save', your Common Data Service environment will be linked to the storage account you provided in earlier step and we will create the file system in the Azure storage account with a folder for each entity you chose to replicate to the data lake.

Under the linked data lake you just created, you can view the status (initial sync status, count of records replicated and last synchronized time stamp) for each of the entities



The screenshot shows the PowerApps interface for 'Export to data lake (preview)'. The left sidebar contains navigation options: Home, Learn, Apps, Create, Data (selected), Entities, Option Sets, Dataflows, Export to data lake (preview) (highlighted), Connections, Custom Connectors, and Gateways. The main area displays a table of entities with columns: Entity, Name, Initial sync status, Last sync status, Last synchronized on, and Count. The table lists four entities: Account, Contact, Lead, and Opportunity, all with 'Completed' initial sync status and 'Success' last sync status.

Entity	Name	Initial sync status	Last sync status	Last synchronized on	Count
Account	account	Completed	Success	09/15/2019 10:44:37 PM	194
Contact	contact	Completed	Success	09/15/2019 10:44:12 PM	216
Lead	lead	Completed	Success	09/15/2019 10:43:45 PM	19
Opportunity	opportunity	Completed	Success	09/15/2019 10:44:35 PM	87

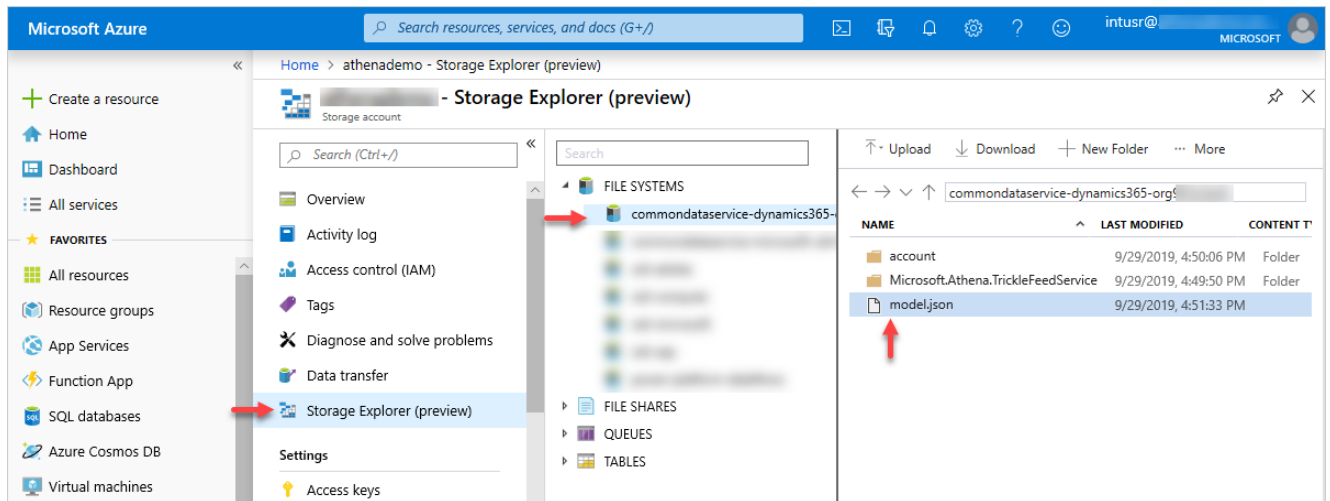
You can use the 'Link to data lake' wizard to link additional data lakes to this environment and 'Unlike data lake' to unlink your environment.

For ongoing administration, use the 'Manage entities' wizard to add/remove entities.

Viewing your data in Azure data lake

The replicated data is store in the Azure data lake in the Common Data Model format. You can view your replicated data in Azure data lake storage by logging into <https://portal.azure.com>. After you login, select the storage account and under Storage Explorer\File System you would see a container with your environment name under which you would see a folder for each of the entities you chose to replicate to the lake along with the model.json file. The metadata file (or model.json) in a Common Data Model folder describes the data in the folder, metadata and location.

Here is an example of the Account entity (stored in the Account folder in csv format) replicated to the lake along with the model.json file



Snip of the model.json file showing the Account entity metadata

```
{
  "name": "cdm",
  "description": "cdm",
  "version": "1.0",
  "entities": [
    {
      "$type": "LocalEntity",
      "name": "account",
      "description": "account",
      "attributes": [
        {
          "name": "Id",
          "dataType": "guid"
        },
        {
          "name": "SinkCreatedOn",
          "dataType": "dateTime"
        },
        {
          "name": "SinkModifiedOn",
          "dataType": "dateTime"
        },
        {
          "name": "statecode",
          "dataType": "int64"
        },
        {
          "name": "statuscode",
          "dataType": "int64"
        },
        {
          "name": "address1_addresstypecode",
          "dataType": "int64"
        },
        {
          "name": "msdyn_travelchargetype",
          "dataType": "int64"
        },
        {
          "name": "address1_shippingmethodcode",
          "dataType": "int64"
        },
        {
          "name": "address1_freighttermscode",
          "dataType": "int64"
        },
        {
          "name": "accountratingcode",
          "dataType": "int64"
        }
      ]
    }
  ]
}
```

Support for initial and incremental writes for data and metadata

Export to data lake service support initial and incremental writes for data and metadata. Any data or metadata changes in the Common Data Service is automatically pushed to the lake without any additional actions. We use the change tracking feature in the Common Data Service to provide a way to keep the data synchronized in an efficient manner by detecting what data has changed since the data was initially extracted or last synchronized.

<incremental metadata example> Snip of how adding a custom field on sources shows-up in model.json

```
{
  "name": "address1_upszone",
  "dataType": "string"
},
{
  "name": "modifiedonbehalfbyyominame",
  "dataType": "string"
},
{
  "name": "parentaccountidyominame",
  "dataType": "string"
},
{
  "name": "address2_primarycontactname",
  "dataType": "string"
},
{
  "name": "opendeals",
  "dataType": "int64"
},
{
  "name": "crd5f_customfield1",
  "dataType": "string"
},
{
  "name": "crd5f_customfield1",
  "dataType": "string"
}
```

Please ensure that your entities have been enabled for change tracking. Please click [here](#) for more details on change tracking.