Search unstructured data in Azure cloud Storage

rogara

# Search unstructured data in cloud storage

In this two-part tutorial series, you learn how to search semi-structured and unstructured data. In this part, you search unstructured data. Unstructured data is data that either is not organized in a pre-defined manner or does not have a data model. An example would be a .txt file.

In this part you learn how to:

[!div class="checklist"] \* Create a resource group \* Create a storage account \* Create a container \* Upload data to your container \* Create a search service through the portal \* Use the search service to search your container

## Download the sample

A sample data set has been prepared for you. **Download** [**clinical-trials.zip**](https://github.com/roygara/storage-blob-integration-with-cdn-search-hdi/raw/master/clinical-trials.zip) and unzip it to its own folder.

Contained in the sample is a set of text files obtained from [clinicaltrials.gov](https://clinicaltrials.gov/ct2/results). You are using them as example text files to search using Azure.

## Log in to Azure

Log in to the [Azure portal](http://portal.azure.com).

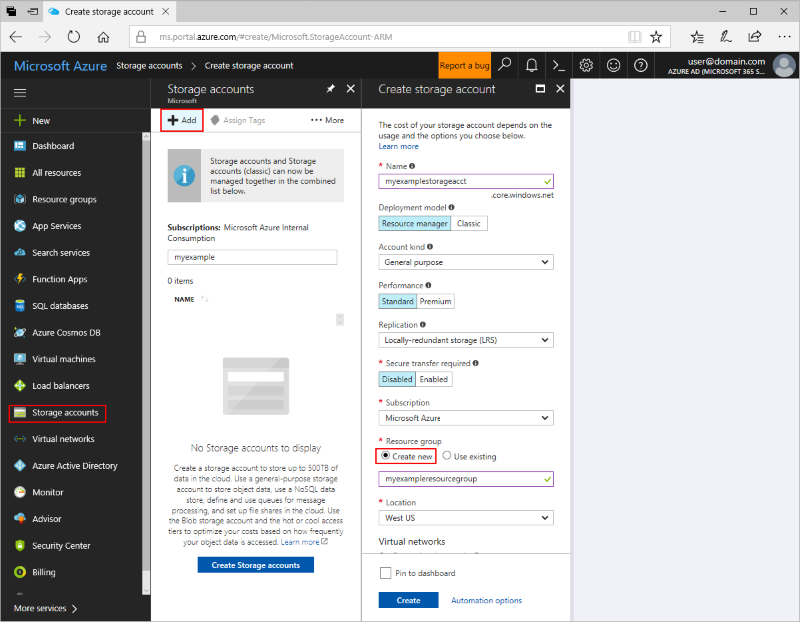
## Create a storage account

A storage account provides a unique location to store and access your Azure Storage data objects.

Currently, there are two types of storage accounts, **Blob** and **General-purpose**. For this tutorial, you create a **General-purpose** storage account.

If you are not familiar with the process of creating a General-purpose storage account, here's how to create one:

1. From the left menu, select **Storage Accounts**, then select **Add.**
2. Enter a name for your storage account.
3. Leave Deployment model as **Resource Manager** and select **General purpose** from the Account kind drop-down.
4. Select **Locally-redundant storage (LRS)** from the Replication drop-down.
5. Under **Resource group**, select **Create new** and enter a unique name.
6. Leave the remaining values as their defaults but be sure to select an appropriate subscription.
7. Choose a Location and select **Create.**



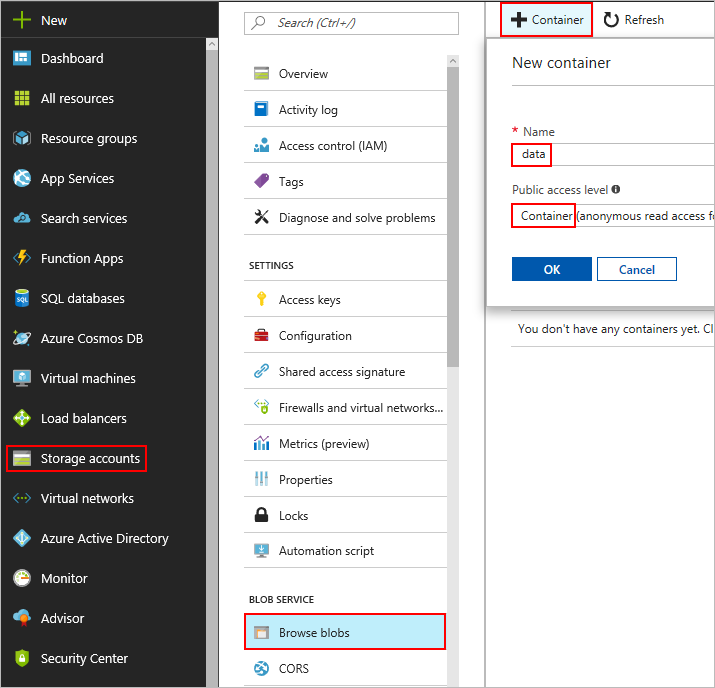
Unstructured search

## Create a container

Containers are similar to folders and are used to store blobs.

For this tutorial, you use a single container to store the text files obtained from clinicaltrials.gov.

1. Navigate to your storage account.
2. Select **Browse blobs** under **Blob Service.**
3. Select **+Container** to create a new container.
4. Name the container "data" and select **Container** for the public access level.
5. Select **OK** to create the container.

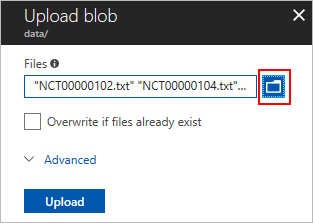


Unstructured search

## Upload the example data

Now that you have a container, you can upload your example data to it.

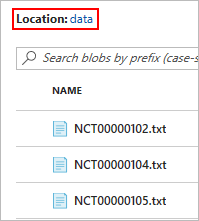
1. Select your container and select Upload.
2. Select the blue folder icon pictured next to the Files field.
3. Navigate to the local folder where you extracted the sample data.
4. Select the extracted files and select **Open**
5. Select **Upload** to begin the upload process.



Unstructured search

The upload process may take a moment.

After it completes, navigate back into your data container, to confirm the text files uploaded.



Unstructured search

## Create a Search Service

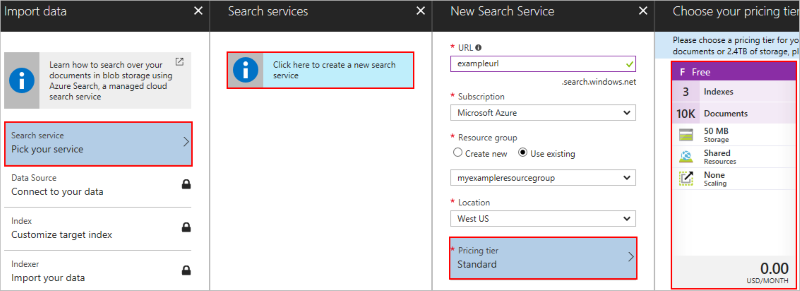
Azure Search is a search-as-a-service cloud solution that gives developers APIs and tools for adding a rich search experience over your data in web, mobile, and enterprise applications.

If you are not familiar with the process of creating a search service, here's how to create one:

1. Navigate to your storage account.
2. Scroll down and click **Add Azure Search** under **BLOB SERVICE.**
3. Under **Import Data**, select **Search Service**.
4. Select **Click here to create a new search service**.

Under **New Search Service**.

1. Enter a unique name for your search service in the **URL** field.
2. Under **Resource group**, select **Use existing** and choose the resource group you created earlier.
3. For the **Pricing tier**, select the **Free** tier and click Select.
4. Select **Create** to create the search service.



Unstructured search

## Connect your search service to your blob storage

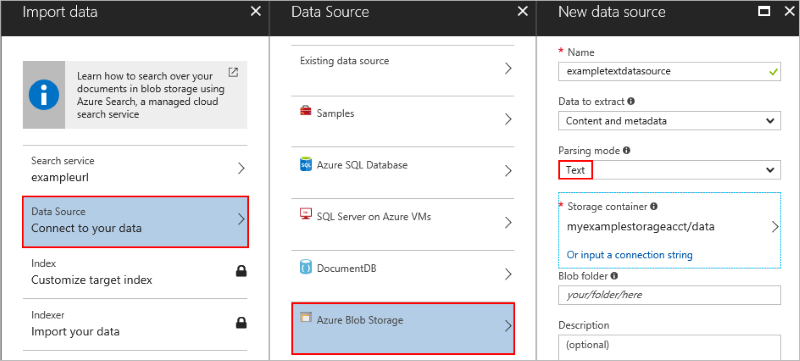
Now that you have a search service, you can attach it to your blob storage. This section walks you through the process of choosing a data source, creating an index, and creating an indexer.

1. Navigate to your storage account.
2. Select **Add Azure Search** under **BLOB SERVICE.**
3. Select **Search Service** inside **Import Data** and then click the search service you created in the preceding section. This opens up **New data source**.

### New Data Source

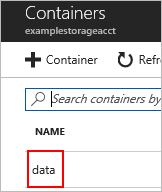
A data source specifies which data to index, how to access the data. A data source can be used multiple times in the same search service.

1. Enter a name for the data source. Under **Data to extract** select **Content and Metadata**. This specifies which parts of the blob are indexed.
   1. In the future, for your own data, you can also select **Storage metadata only** if you wish to limit the data that is indexed to standard blob properties or user-defined properties.
   2. You can also choose **All metadata** to obtain both standard blob properties and *all* content-type specific metadata.
2. Since the blobs you're using are text files, set **Parsing Mode** to **Text**.



Unstructured search

1. Select **Storage Container** to list the available storage accounts.
2. Select your storage account then select the container you created in the preceding step.
3. Click **Select** to return to **New data source** and select **OK** to continue.

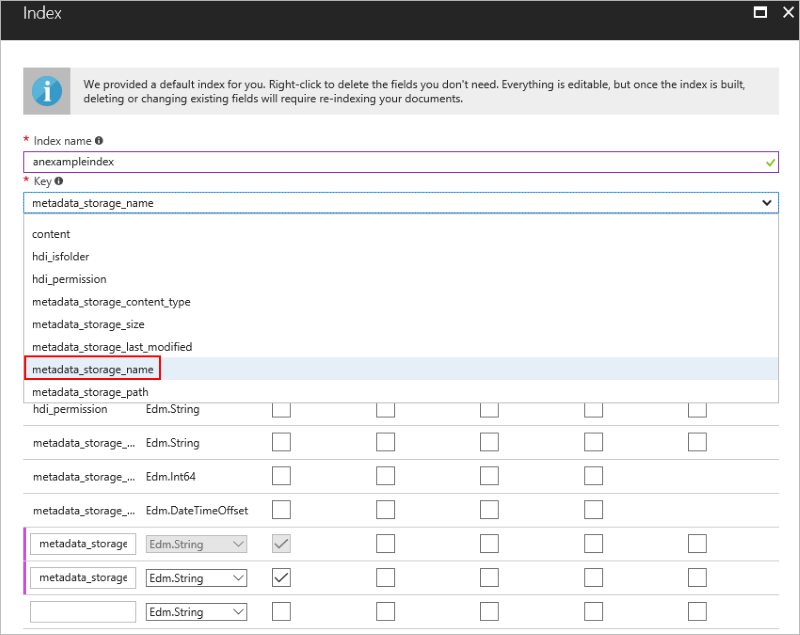


Unstructured search

### Configure the Index

An index is a collection of fields from your data source that are able to be searched.

1. Enter a name for your index in the **index name** field.
2. Select **metadata\_storage\_name** from the **Key** dropdown.



Unstructured search

1. Click **OK**, which brings up **Create an Indexer.**

The parameters of your index and what attributes you give those parameters are important. The parameters specify *what* data to store, the attributes specify *how* to store that data.

The following table provides a listing of the available attributes and their descriptions.

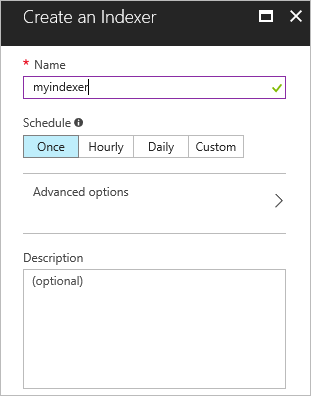
### Field attributes

|  |  |
| --- | --- |
| Attribute | Description |
| *Key* | A string that provides the unique ID of each document, used for document lookup. Every index must have one key. Only one field can be the key, and its type must be set to Edm.String. |
| *Retrievable* | Specifies whether a field can be returned in a search result. |
| *Filterable* | Allows the field to be used in filter queries. |
| *Sortable* | Allows a query to sort search results using this field. |
| *Facetable* | Allows a field to be used in a faceted navigation structure for user self-directed filtering. Typically fields containing repetitive values that you can use to group multiple documents together (for example, multiple documents that fall under a single brand or service category) work best as facets. |
| *Searchable* | Marks the field as full-text searchable. |

### Create an Indexer

An indexer connects a data source with a search index, and provides a schedule to reindex your data.

1. Enter a name in the **Name** field and select **OK**.



Unstructured search

1. You are brought back to **Import Data** select **OK** to complete the connection process.

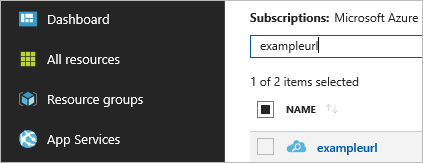
You've now successfully connected your blob to your search service. The search service begins indexing immediately and you can begin searching right away.

## Search your text files

To search your files, you need to open the search explorer inside the index of your newly created search service.

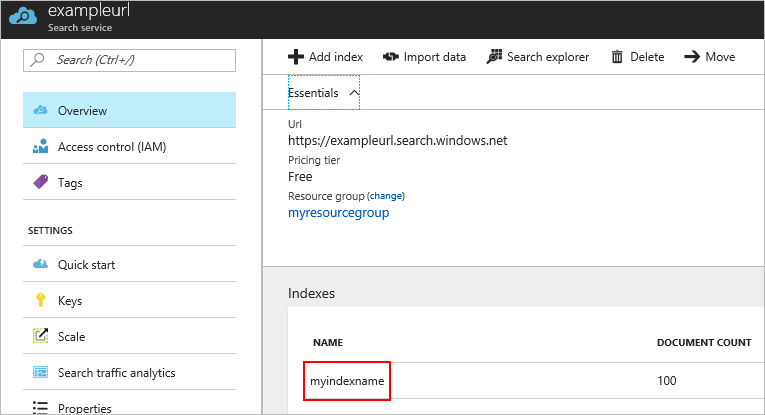
The following steps show you where to find it and provide you some example queries:

1. Navigate to all resources and find your newly created search service.



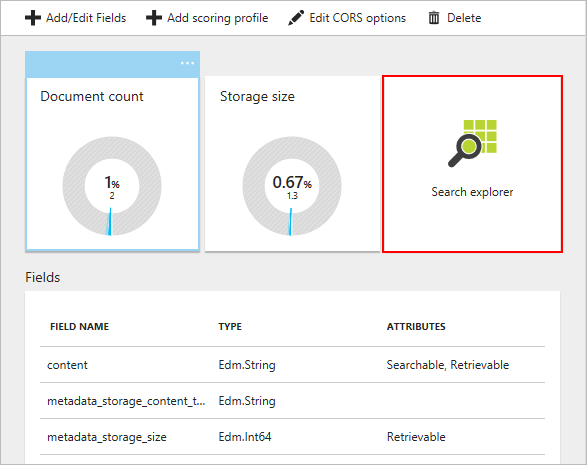
Unstructured search

1. Select your index inside of indexes to open it up.



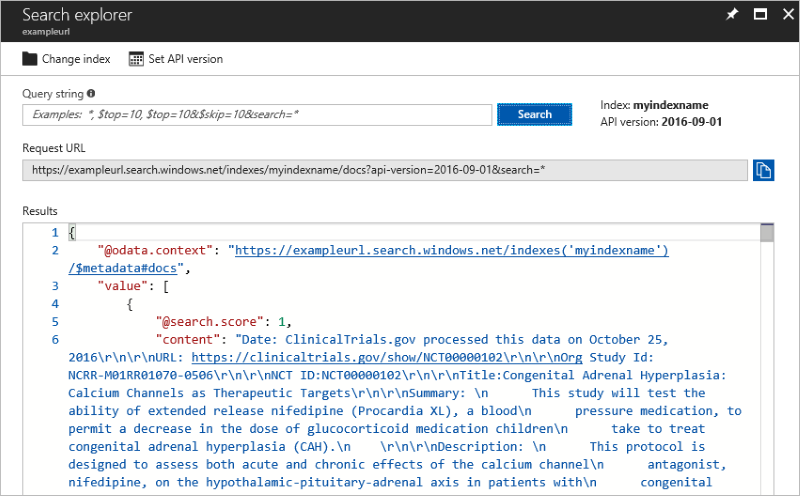
Unstructured search

1. Select **Search Explorer** to open up the search explorer, where you can make live queries on your data.



Unstructured search

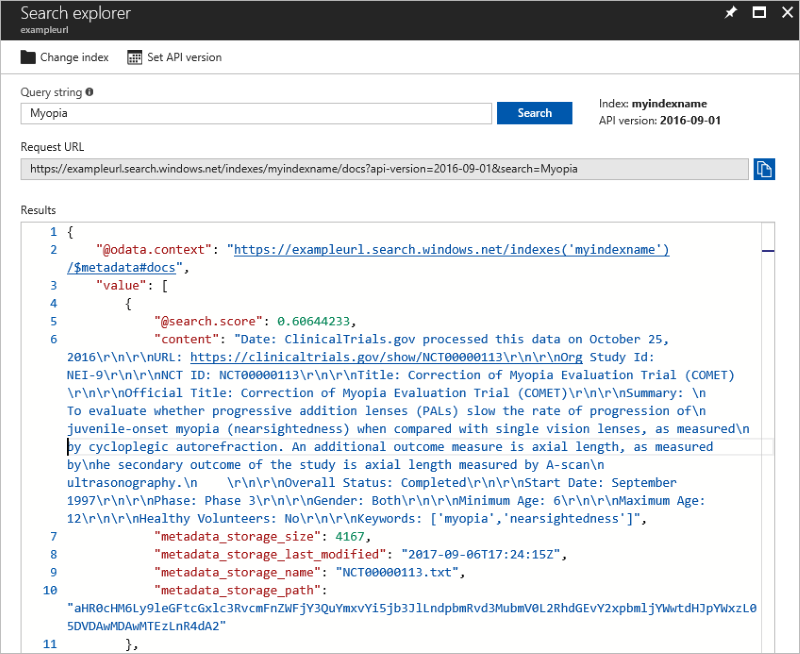
1. Select **Search** while the query string field is empty. This returns *all* the data from your blobs.



Unstructured search

### Full-text search

In the **Query string** field "Myopia" and select **Search**. This initiates a search of the contents of the files, returns a subset of them that contain the word "Myopia."



Unstructured search

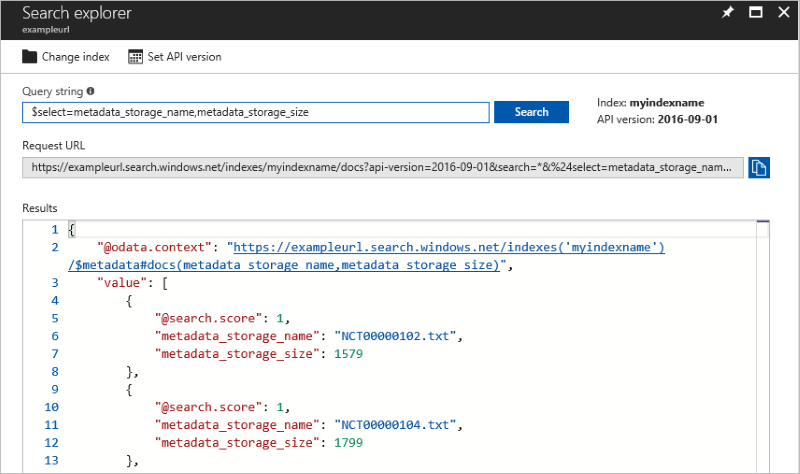
### System properties search

You can also create queries that search by system properties using the $select parameter.

Enter $select=metadata\_storage\_name into the query string and press enter, returning only that particular field.

The query string is directly modifying the URL, so spaces are not permitted. To search multiple fields, use a comma, such as: $select=metadata\_storage\_name,metadata\_storage\_size

The $select parameter can only be used with fields that are marked retrievable when defining your index.



Unstructured search

You have now completed part one of this series and have a searchable set of unstructured data.

Make sure **not** to delete any resources created in this tutorial, as most of them are going to be reused in the subsequent tutorial.

## Next steps

In this tutorial, you learned about searching unstructured data such as how to:

[!div class="checklist"] \* Create a resource group \* Create a storage account \* Create a container \* Uploading data to your container \* Create a Search Service \* Using the Search Service to search your container

Advance to the next tutorial to learn about semi-structured search.

[!div class="nextstepaction"][Searching Structured Data](./storage-semi-structured-search.md)