Creating a Search Index in the Azure Portal

The steps below will walk you through creating a search index in the Azure Portal using the import data wizard. The web app you'll spin up in the next step expects to receive several fields to work properly. Because of this, be sure to follow the requirements outlined below to avoid any problems during setup.

Requirements

If you choose to create your index via the Azure Portal, set your search field properties as described in the table below:

Field	Notes and Expectations
content	Used to show the transcript of the files. Should be **searchable and retrievable**
metadata_storage_path	This should be the **key field**. The storage path is used to query the blob indexer for the content so that you can "preview" the file. Should be **retrievable**. If it is **base64 encoded**, make sure to specify it in the appsettings.json file of the front end application.
metadata_storage_name	The storage name is used to display the name of the file on the results page. Should be **retrievable**.
people	List of strings with the persons identified in the document. Should be **facetable, filterable, searchable and retrievable**.
locations	List of strings with the locations identified in the document. Should be **facetable, filterable, searchable and retrievable**.
organizations	List of strings with the organizations identified in the document. Should be **facetable, filterable, searchable and retrievable**.
keyPhrases	List of strings with the key phrases identified in the document. Should be **facetable, filterable, searchable and retrievable**.
language	String containing the main language of the document. Should be **facetable, filterable, searchable and retrievable**.

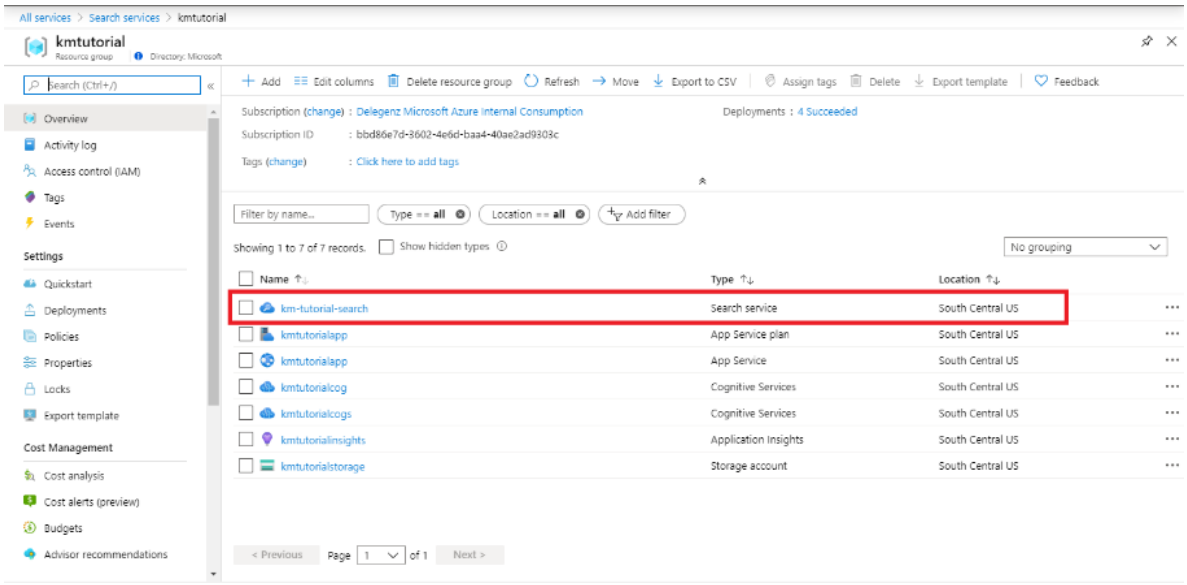
Instructions

Follow the steps and screenshots below to create your index.

Exercise 1.- Create a Search Index Using the Portal

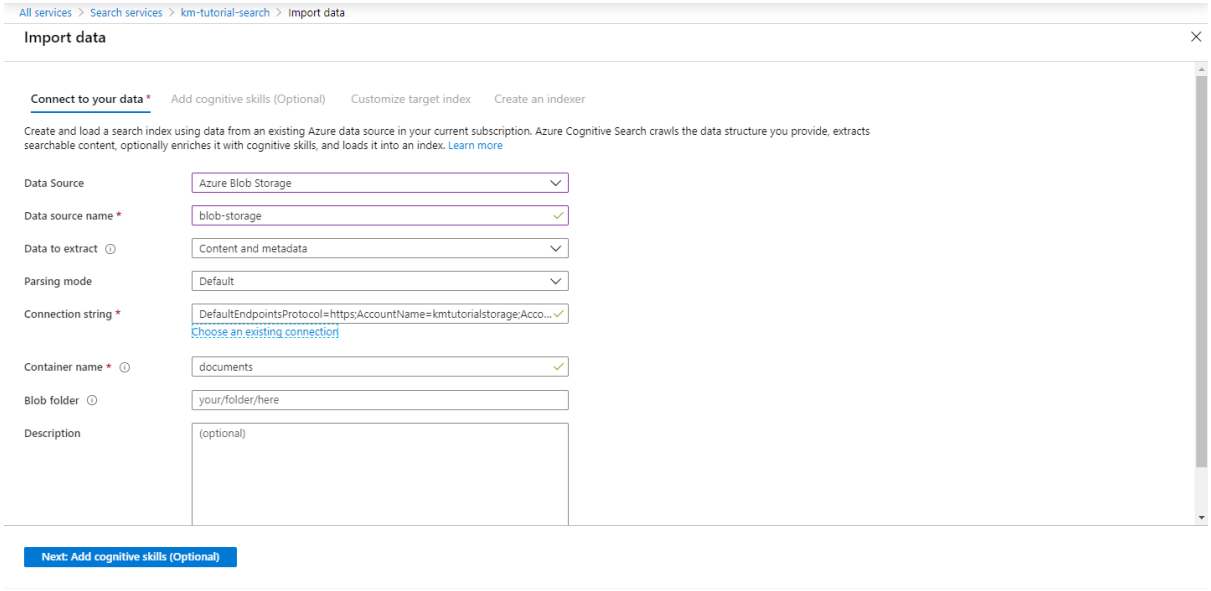
Task 1. Navigate to your Search Service

1. Open your browse and navigate to www.portal.azure.com.
2. From the main menu, select under the favourites menu the **Resource Groups**.
3. From the resource group blade select click on the resource group available.
4. Locate the Search Service called **xxx-search-service**.



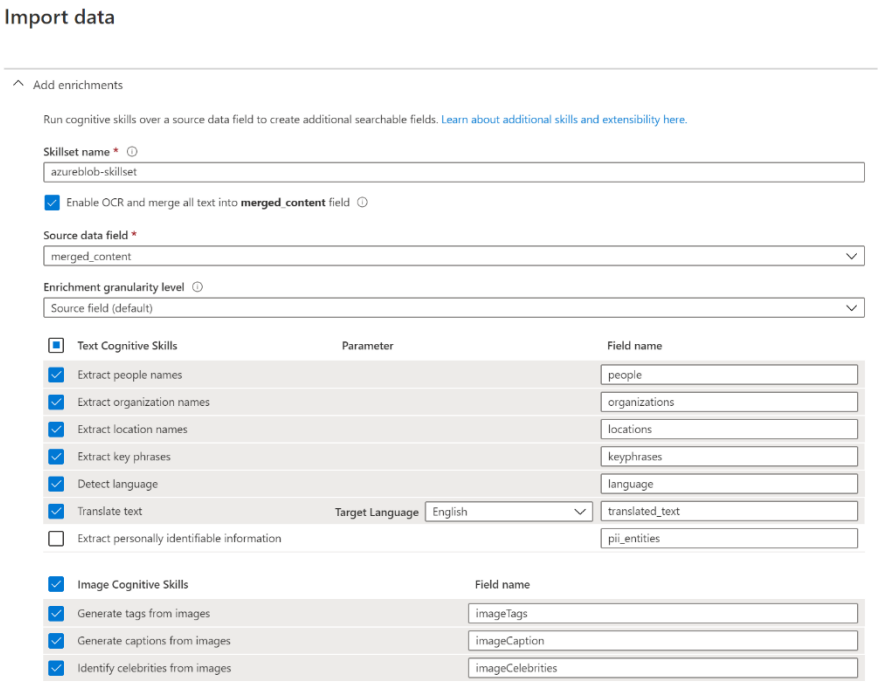
Task 2. Import data

1. In the Search Service blade, on the top menu select the **Import Data** tab.
2. Inside the Import Data blade, make sure you are located on the **Connect to your Data** tab.
3. From the **Connect to your Data** blade, select the **Data Type** option and from the dropdown select **Azure Storage Explorer.**
4. Keep the defaults and use **Choose an existing connection** to connect to your storage account and select the **xxx-data** container with your data.
5. At the bottom of the page, click **Next: Add cognitive skills (Optional)**



Task 3. Add cognitive skills

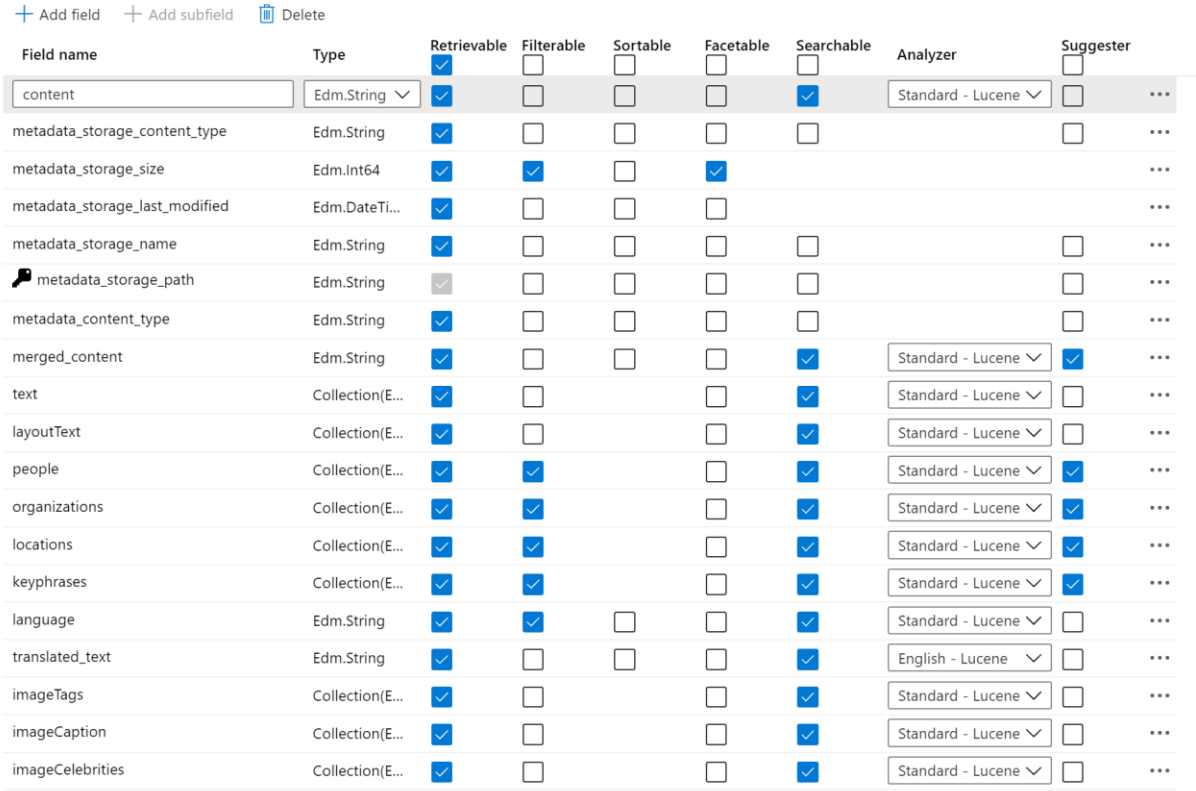
1. From the **Add cognitive skills (Optional)** tab, first select and expand the **Attach Cognitive Services** menu.
2. A Cognitive Services resource should have been automatically provisioned, just select it. If it has not been provisioned successfully, use the **Refresh** button.
3. Collapse the **Attach Cognitive Services** menu.
4. Select and expand the **Add Enrichments** menu.
5. Replace the name of the skillset using your initials xxx-skillset.
6. Check the option for enabling **OCR** located below the skillset name and leave default settings.
7. Check the option for **Text Cognitive Skills**.
8. Please note that the **Extract personally identifiable information** skill currently does only support English documents. Some of the sample data is currently in different languages, hence do not enable it when you use the sample documents from this lab.
9. Check the option for **Image Cognitive Skills**.



10. Collapse the **Add Enrichments** menu.
11. Select and expand the **Save enrichments to knowledge store (Preview)**.
12. Below the **Storage Account connection string** box, click on the option **Choose and existing connection** and select your storage account **xxxstr**.
13. Once selected check the option **Azure Table projections**.
14. Collapse the **Save enrichments to knowledge store (Preview)** menu.
15. At the bottom of the page, click **Next: Customize target index**.

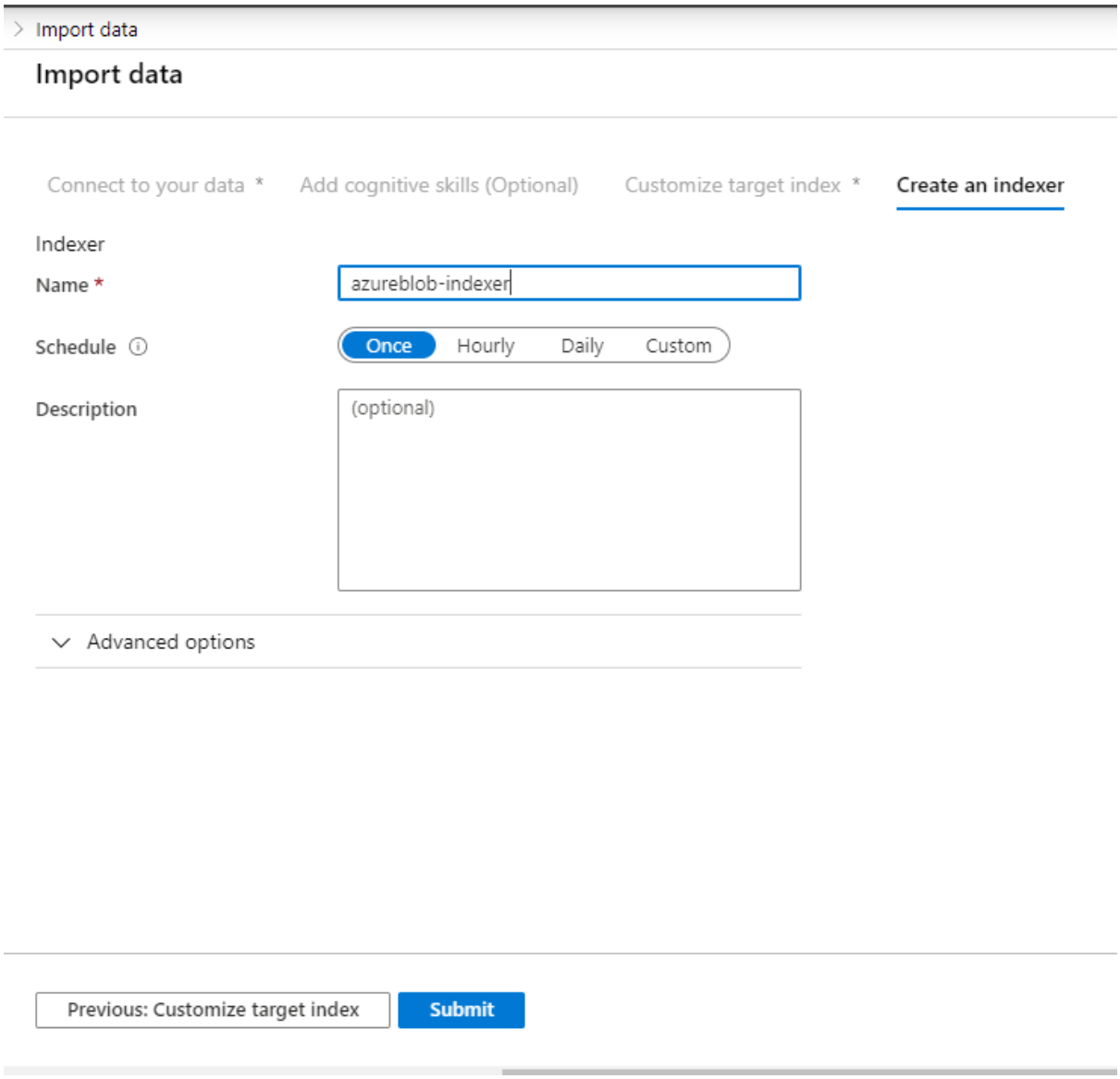
Task 4. Customize target index

1. From the **Customize target index** blade, replace the index name with your initials **xxx-index**.
2. Make sure that the value inside the Key box is **metadata_storage_path**.
3. Name the **Suggester** using your initial **xxx-suggester**.
4. Set the **Search Mode** to **analyzingInfixMatching**. The Suggester feature provides type-ahead suggestions, as you can see in web search engines like Bing.
5. Set your configuration like the image below and then select at the bottom **Next: Create an indexer button**. A validation will be made.



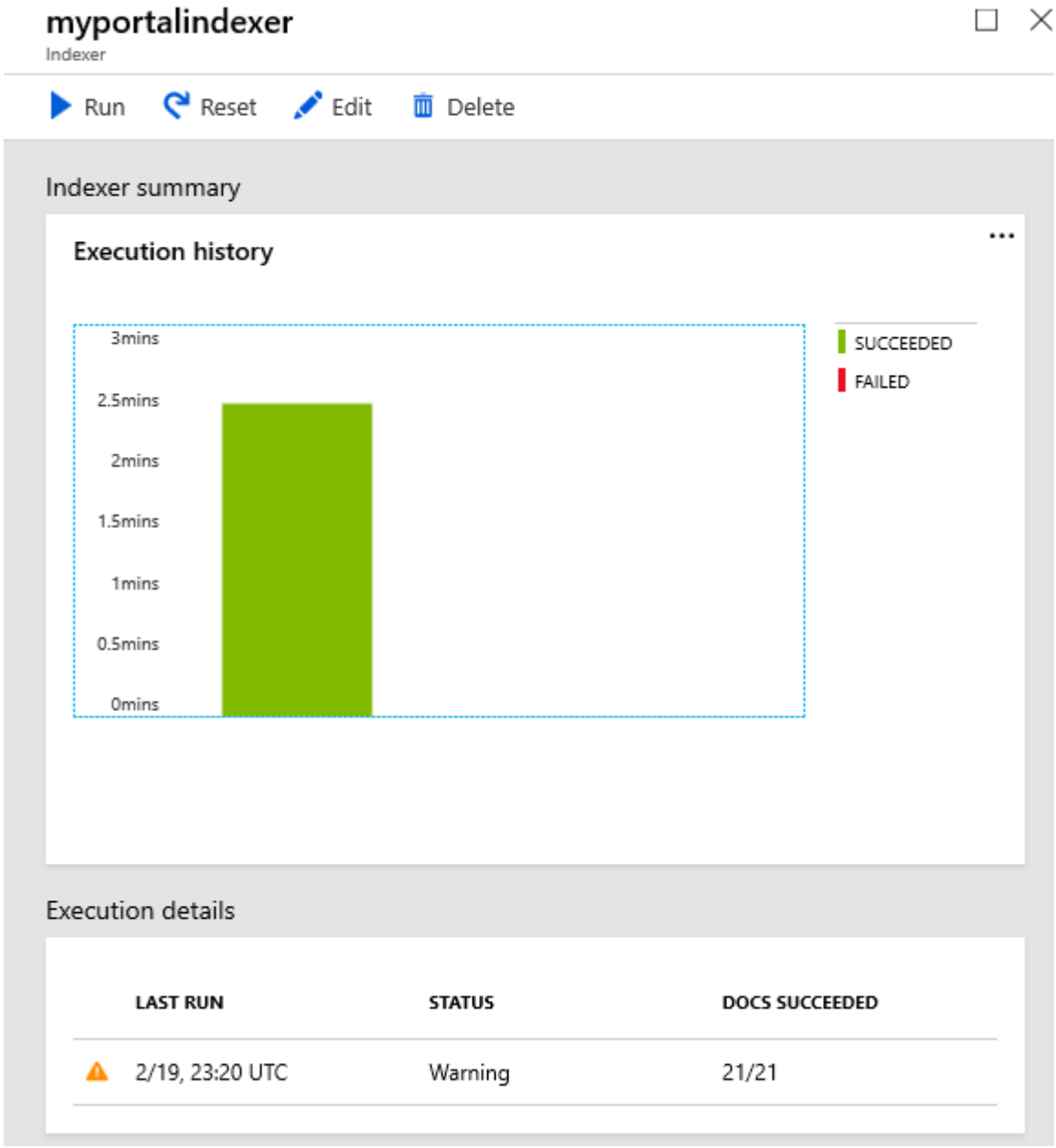
Task 5. Create an Indexer

1. Name your indexer as **xxx-indexer**. The indexer is the job that connects the data source, the index and the schedule.
2. Set the schedule as once.
3. Expand the Advanced Options link.
4. Set **Max failed items** to **-1**, we don't want the indexer to stop processing a document even when any cognitive skill has an error.
5. Set **Max failed items per batch** to `-1`, we don't want the indexer batch job to stop at any reason.
6. Set **Data to extract** to `Content and metadata`, since we are using both as you can see above.
7. Set **Parsing mode** to **Default**, we have both text and image skills
8. Select the blue Submit button, and you will be redirected to the overview tab, where now you can see 1 index, 1 indexer and 1 data source (you may have to refresh your page).
9. Select the Indexes, Indexers, Data sources, and Skillsets tabs to see the objects you just created using the portal. In the next lab, you will learn how to create them with REST API calls.



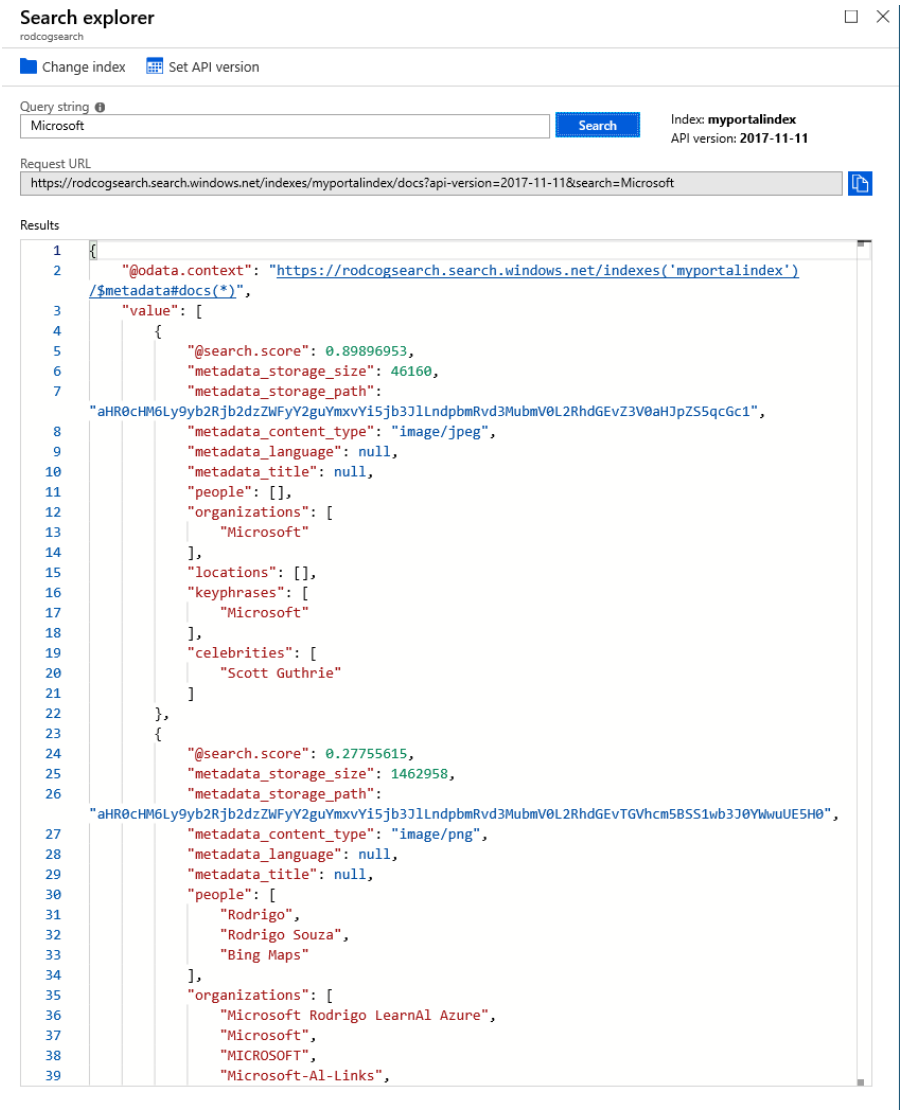
Task 6. Check Indexer Execution Status

1. To monitor data import, select the Indexers link (it is in the middle of the page and you can also see it in the middle of the image above). It should be In progress. Wait until you get the Warning status, it is expected to have some warnings.
2. You should see the newly created indexer in the list, with status indicating “In progress”, “Failed”, or “Warning”. If not, select the refresh button in the top-middle of the overview tab. The final expected status is “Warning”, along with the number of documents indexed. Warnings are caused by extra long words and big texts. The indexer knows how to deal with them, but warns you.
3. Select the refresh button, top middle of the page, until the execution is over. The “Warning” status is expected, select the Indexer name to see the summary. In this page you will see all the executions this Indexer may have and its details, duration and so on.



Task 7. Query the Azure Cognitive Search Index

1. Navigate back to the **Overview Tab** of the Azure Cognitive Search service and select the **Indexes** link, the second from the left to the right. You should see the "Document Count" and the Storage Size. The expected count is 20 and the size should be close to 10.89 MiB. As you can see, Azure Cognitive Search doesn't store all the document contents, just parts of them: key words, metadata, tags.
2. From the top, select the **Search Explorer** link.
3. Select **Search** to search for all documents. You can use any valid simple or full Lucene query syntax to create the request. The * character is equivalent to an empty or unspecified search that returns all documents in no particular order.
4. Try searching for "Microsoft", a different result set is expected. In the resulting json, you'll see a number after `@search.score`. Scoring refers to the computation of a search score for every item returned in search results.
5. You will probably see results like the image below. Scroll down until you see all meta information available.



Exercise 2.- Visualizing the Results with a Demo FrontEnd

Task 1. Start the VS project

1. Clone the repository from the following address:
   git clone https://github.com/Azure-Samples/azure-search-knowledge-mining.git
2. Open CognitiveSearch.UI.csproj 02-Web UI Template\CognitiveSearch.UI) in Visual Studio.

Task 2. Update appsettings Json

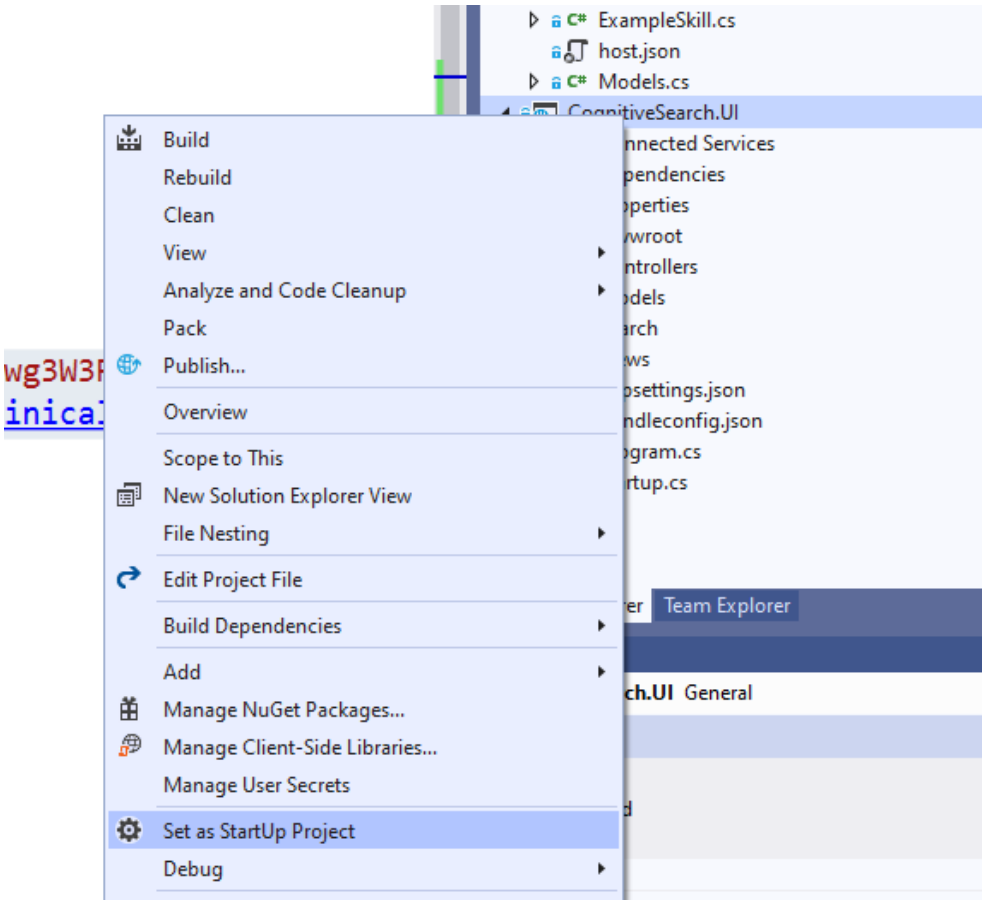
1. Navigate to the following directory: azure-search-knowledge-mining-master\02 - Web UI Template\CognitiveSearch.UI and locate the appsettings.json and replace the values.
2. Update the following fields in the appsettings.json file to connect the web app to your storage account, search index, and app insights account:

```
"SearchServiceName": "Your Search Service Name",
"SearchApiKey": "Your Search Service key",
"SearchIndexName": "clinical-trials-small",
"InstrumentationKey": "",
"StorageAccountName": "Your storage Account Name",
"StorageAccountKey": "Your Storage Account Key",
"StorageContainerAddress": "Your Storage Container Address",
"KeyField": "metadata_storage_path",
"IsPathBase64Encoded": true,
"GraphFacet": "diseases"
```

Notes

- SearchServiceName should be set to the name of the search service. (i.e. "myservice")
- SearchApiKey should be to the name of the search service. (i.e. "B8365AC95521089B7E3FA4CC98435")
- SearchIndexName should be set to the name of the index (i.e. "clinical-trials-small")
- StorageAccountName should be set to the name of the storage account (i.e. "mystorageaccount")
- StorageContainerAddress should be in the following format: "https://storageaccountname.blob.core.windows.net/containername"
- InstrumentationKey is an optional field. The instrumentation key connects the web app to Application Insights in order to populate the Power BI reports.
- KeyField should be set to the field specified as a key document Id in the index. (i.e. "metadata_storage_path")
- Sometimes metadata_storage_path is the key, and it gets base64 encoded. In that case set IsPathBase64Encoded to true.
- The GraphFacet is used for generating the relationship graph, set it to the name of the facet that you would like to use (i.e. "diseases"). Or leave blank if you won't use the node graph.

3. Right click on the CognitiveSearch.UI object inside the object explorer and select the **Set as Startup Project** option.



4. Run the project and see the results.