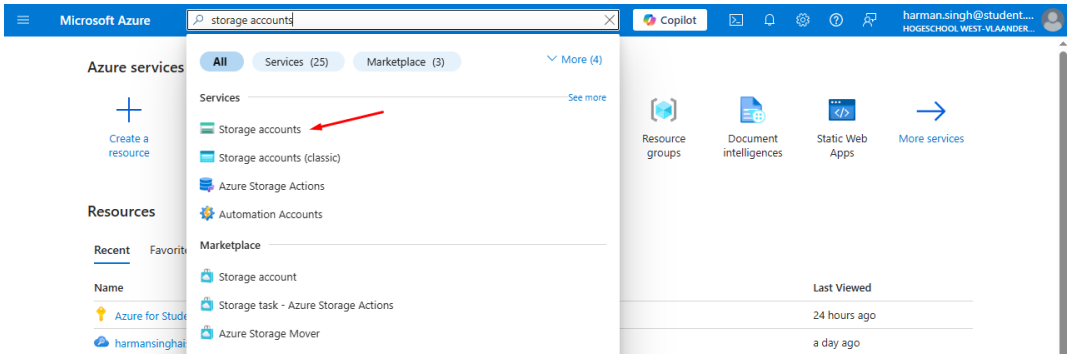
How to (re)build your knowledge base (search index)

In this document you will find the guide to converting your historical data, currently in Excel-format, into data usable by our AI model to generate estimations. For this, you will need to set a few things up, which we will explain here.

Part 1/3: storage account

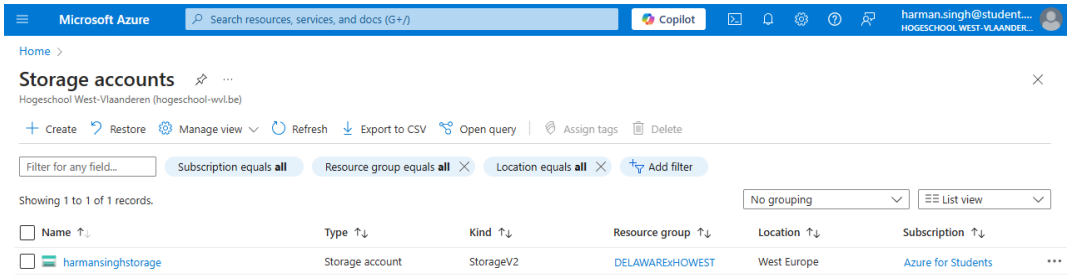
You will need to store your historical data somewhere we can access it. We will create a storage account and upload our files on there. Get started by navigating to the Azure Portal (https://portal.azure.com), and log into your account.

1. Search for “Storage accounts” in the search bar and click on it.



2. After this, you can either make a new storage account or use one that already exists.

   If you chose to make a new one, press the button that says `+ Create` to create a new storage account.

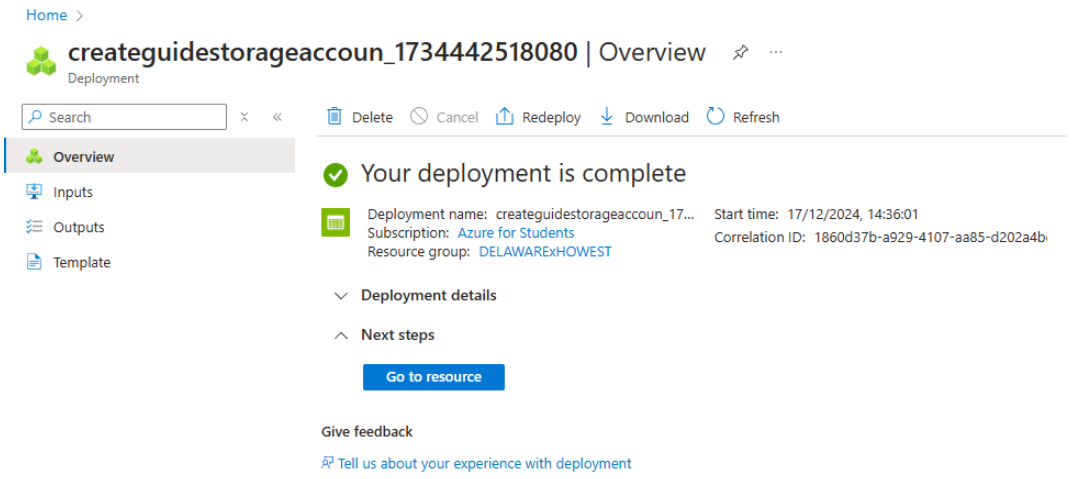


3. This will bring you to a screen where you need to fill in some details about your storage account. Here is how you proceed:

- `Subscription`: select the subscription you want to use in the drop-down list
- `Resource group`: select the resource group you want to add the storage account to
- `Storage account name`: choose a name for the storage account
- `Region`: choose your region, or leave as default
- `Primary service`: don't change
- `Performance`: don't change unless necessary
- `Redundancy`: depends on the needs of your company

You can further configure the resource if you know what you're doing, but that is optional in this case.

4. After this press the blue `Review + create` to create the resource, and then the blue `Create` button on the next screen to actually make the resource.
5. Wait for the deployment to complete. Once it's done, you should see this:



6. Press the blue button that reads `Go to resource`

Part 2/3: containers

To store files on the storage account, we need to create a container. This is very easy to do.

1. In the sidebar, under `> Data storage` press `Containers`.
2. Press the button that reads `+ Container`
3. Enter the name `knowlegde-base`

> 💡 You can also use a different name, but `knowlegde-base` is descriptive. This name will be an environment variable.

4. Enter the container, and press the `Upload` button. Proceeded with `Browse for files`
5. Upload all files you want to add to the knowledge base. These files will be previously (manually-)made estimations. **Tip**: you can select multiple files at once
6. Press the `Upload` button.

If you did everything correctly you should now see your files in the container

Part 3/3: the script

This step is also quite easy to do. You need to run a Python script. This requires your computer to have Python, and the necessary libraries installed. If you don't have Python installed, you can let someone that has Python installed do this Part of the guide

1. Install Python if you have not already from www.python.org/downloads
2. We assume you have the code base, which contains the script, so simply open a Command Prompt, or Powershell (or whatever MacOS uses), and navigate to where you have the codebase on your device

You should be in this path now: `\...\...\DELAWAREXHOWEST>`

3. Type and execute the following command to install all necessary libraries: `pip install -r .\requirements.txt`. If everything went well, you should not see any errors
4. Type and execute the following command to change your directory to the `/scripts/` directory: `cd scripts`. Your path should now read `\...\...\DELAWAREXHOWEST\scripts>`
5. Type and execute the following command to build your knowledge base: `python build_knowledge_base.py`

If everything went well, the very last line of the output should read: **All Excel files have been processed and uploaded.** This should be the case if:

1. Your files all have the correct format (the file should contain a table with the following columns, in no particular order:
   - MSCW
   - Area
   - Module
   - Feature
   - Task
   - Profile
   - MinDays
   - RealDays
   - MaxDays
   - Contingency
   - EstimatedDays
   - EstimatedPrice
   - potential_issues

   We did not come up with these column names. They are the same as the Excel-document we received showing how you currently are making estimations.
2. If you have the correct environment variables set up in a `.env` file in the root of the project. It should contain the following keys:

```
AZURE_STORAGE_CONNECTION_STRING = "placeholder"
# found in resource: Storage account
# path: Security + networking > Access keys > Connection string


AZURE_KNOWLEDGE_BASE_CONTAINER_NAME = "placeholder"
# the name you gave the container inside the storage
# account (we called it `knowledge-base`)


AZURE_SEARCH_ENDPOINT = "placeholder"
# found in the resource: AI Search
# path: Overview > Url
# should end on search.windows.net)
```

```
AZURE_SEARCH_API_KEY = "placeholder"
# found in resource: AI Search
# path: Settings > Keys > Primary/Secondary admin key (pick one)


AZURE_SEARCH_INDEX_NAME = "placeholder"
# the name you gave the index
# If you used the Python script above, it is `tasks-index-excel`)
# you can find it in the `search_index_configuration.json` file
```