Azure Batch lab

Exercise 1. Running parallel image processing

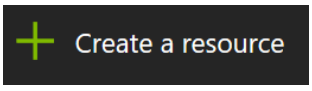
Duration: 45 minutes

In this exercise we will execute a simple ImageMagick resizing job on a set of input jpg files processed in parallel using Azure Batch service and Batch Explorer interface.

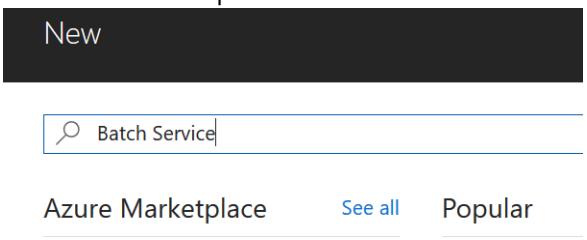
We will use Batch Explorer which is an open-source GUI application for managing Azure Batch accounts, designed to make the use of Azure Batch service easy and straight-forward.

Preparation step 1: Provision Azure Batch account

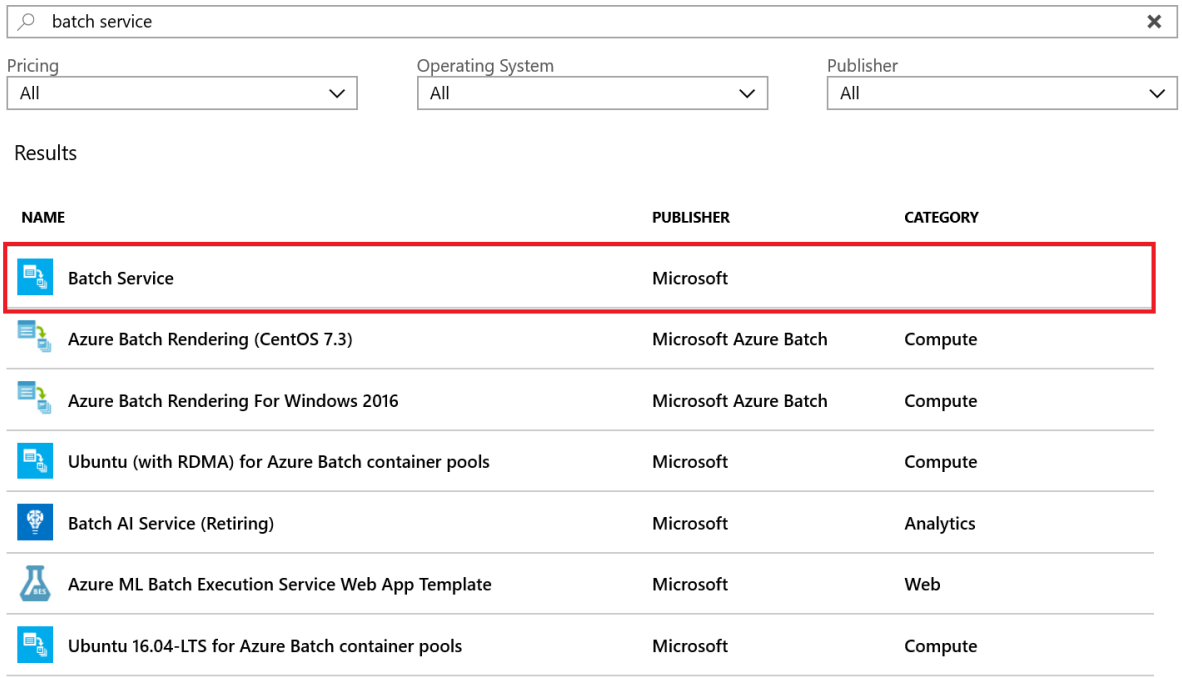
1. Navigate to Azure Portal in the browser.
2. Select **+ Create a resource**.



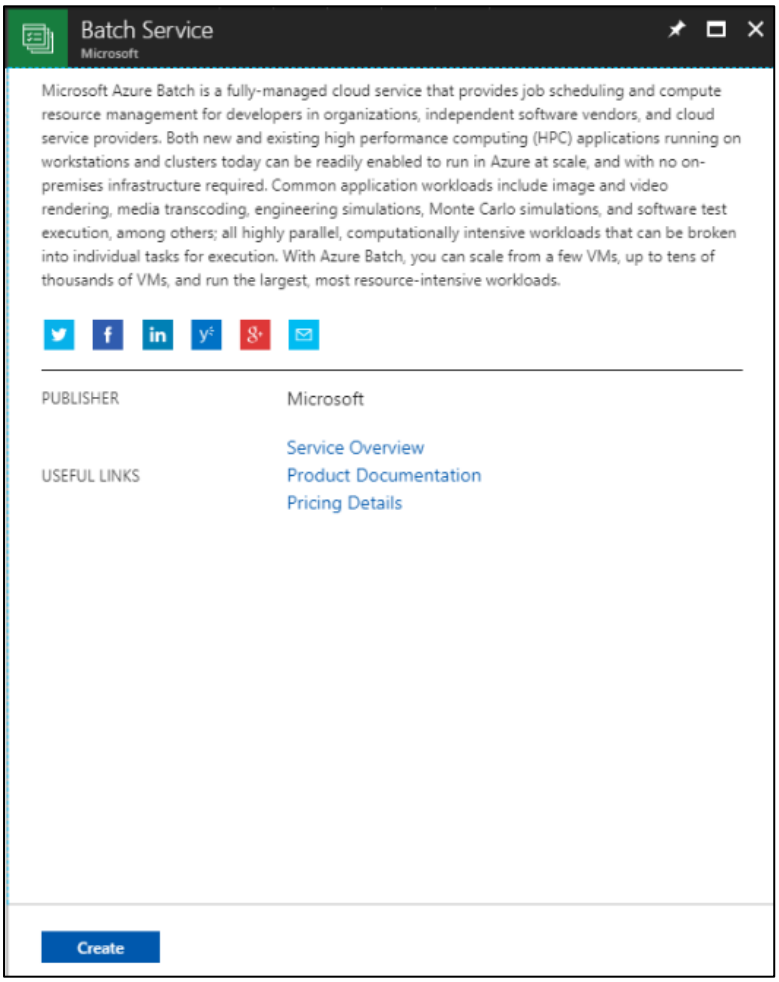
3. Search the Marketplace for **Batch Service**.



4. In the list, select **Batch Service**.

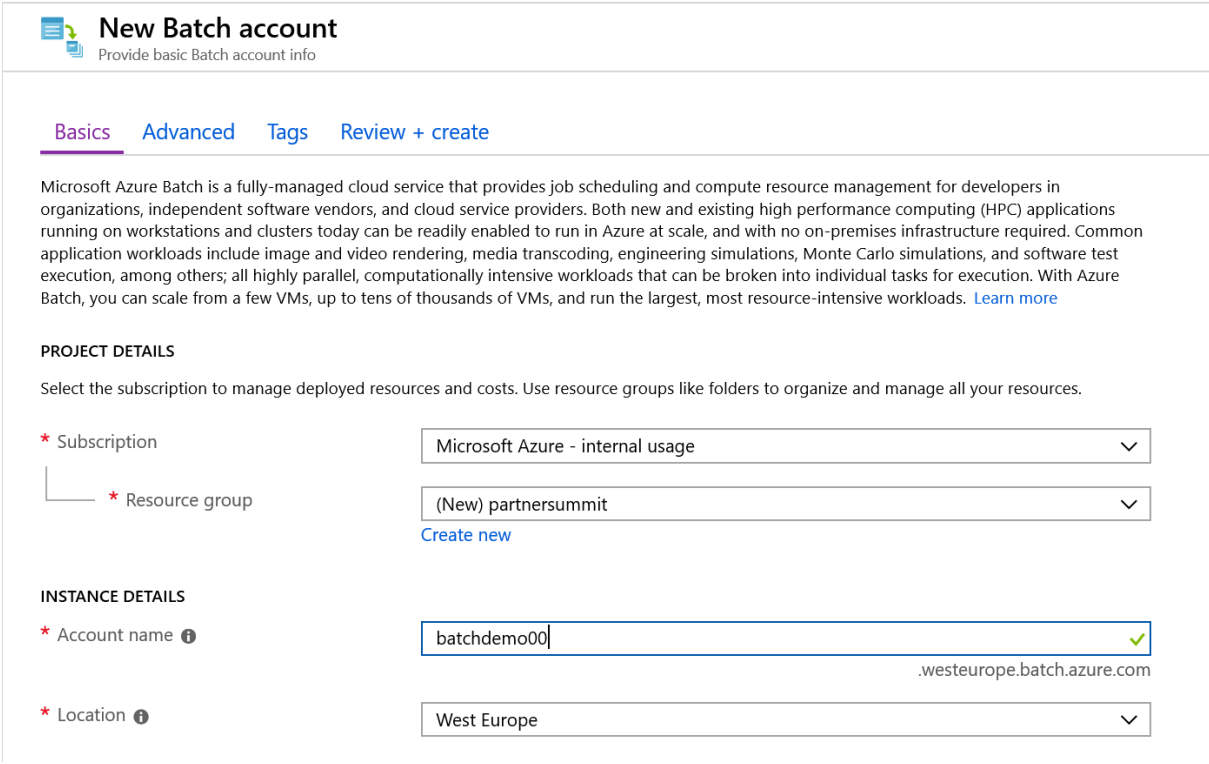


5. Select **Create** on the Batch Service blade.

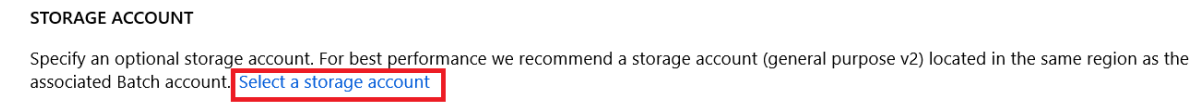


6. On the New Batch Account blade, specify the following:
    a. **Account name**: Provide a name for your new Batch Account. The name you choose must be unique within the Azure region where the account is created (see Location below).
    b. **Subscription**: Select the subscription in which to create the Batch account.
    c. **Resource group**: Select a new resource group and name it (e.g. partnersummit).

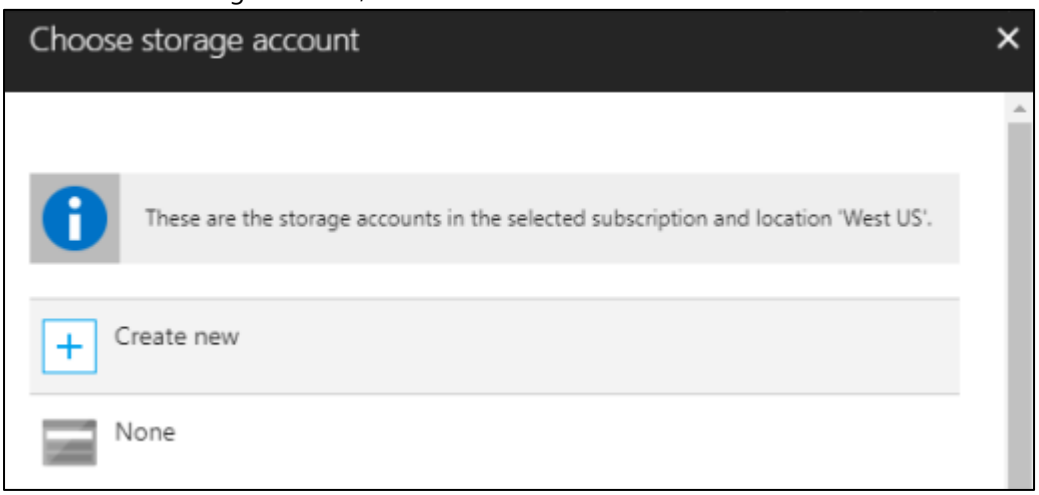
d. **Location**: The Azure region in which to create the Batch account.



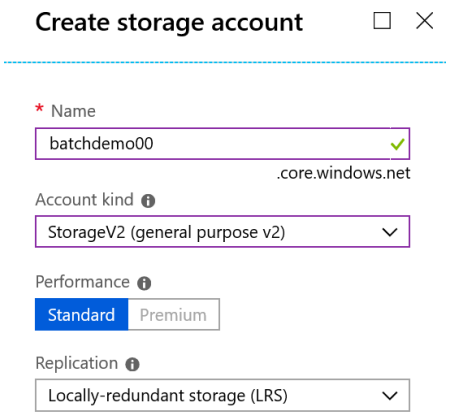
7. Select **Storage account**.



8. In the Choose storage account, select **Create new**.



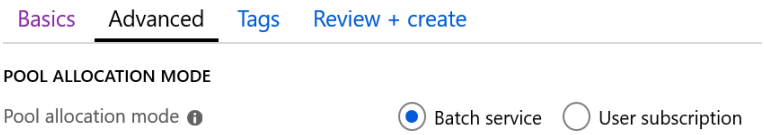
9. On the Create storage account blade:
    a. **Name**: Provide a unique name for the new Azure Storage account that will attached to your Batch Service. E.g. use batchdemo<n> where n is the number on your desk.
    b. **Account kind:** Select StorageV2
    c. **Performance**: Leave this at Standard.
    d. **Replication**: Leave this at Locally-redundant storage (LRS).



e. Select **OK** at the bottom of the blade.

10. Click "Next: Advanced" button at the bottom of the page:

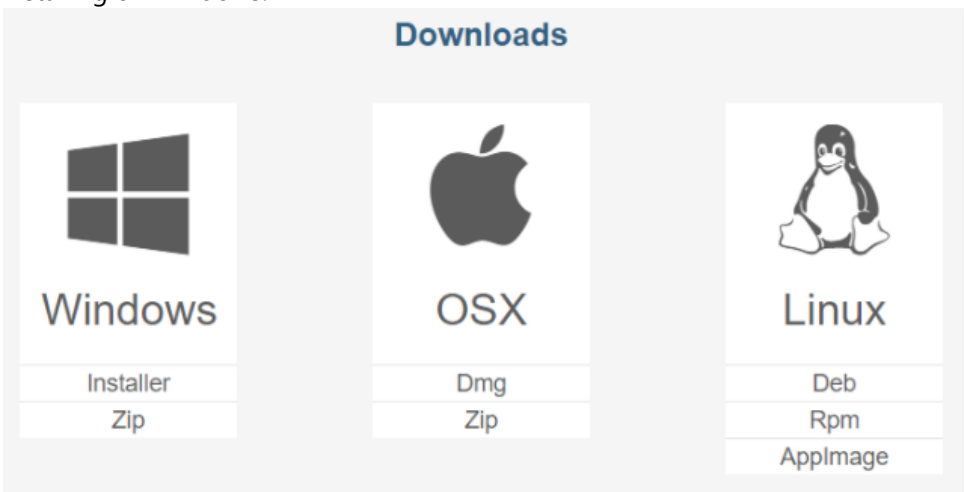
11. Make sure that Pool allocation mode is set to **Batch service**.



12. Click **Review+Create** and then **Create** button to create the new Batch account. The provisioning should take around 1 minute.

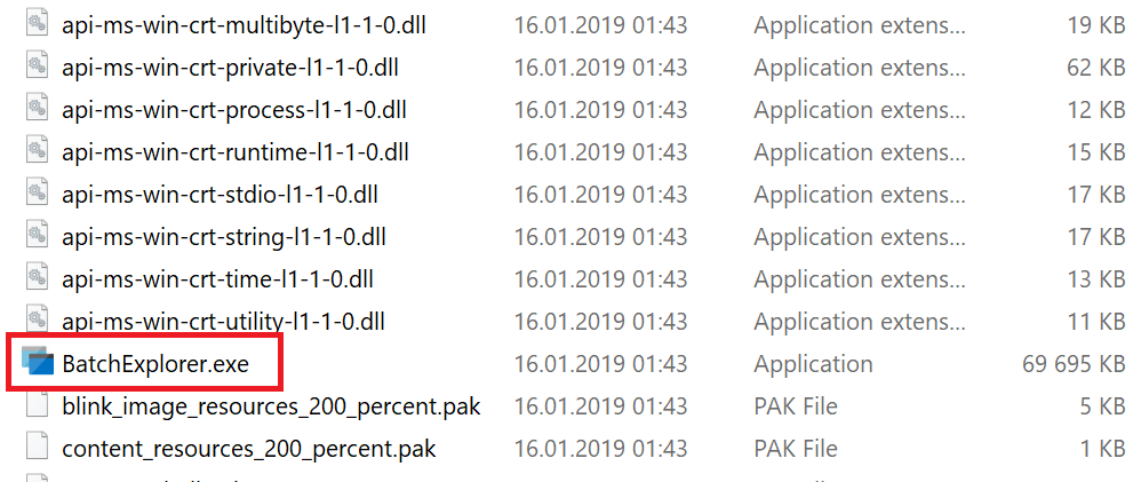
Preparation step 2: Download Batch Explorer

1. Using a web browser on your local machine, navigate to https://azure.github.io/BatchExplorer/.
2. Scroll down until you see the **Downloads** section.
3. Select the Zip download appropriate to your OS. The instructions that follow assume you are installing on Windows.

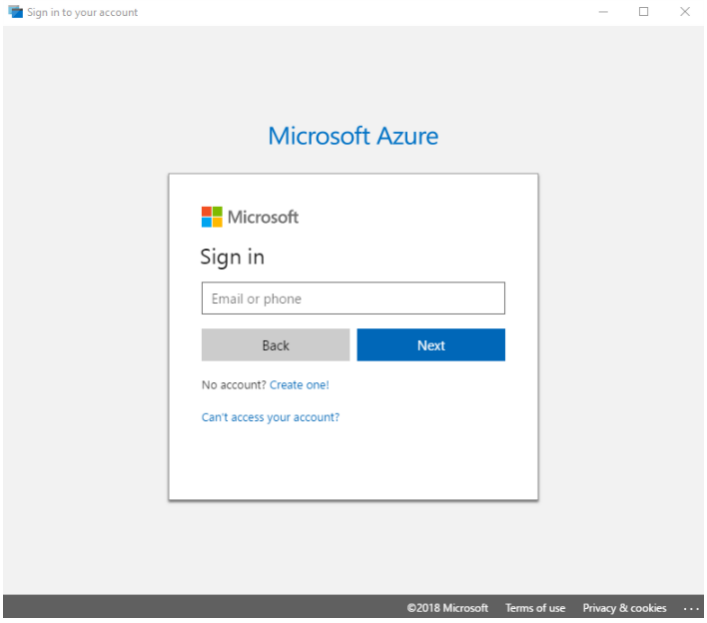


Further instructions assume that **Batch Explorer 0.19.1 (beta)** is installed.

4. Unzip the installer file to a local folder and run the Batch Explorer executable:

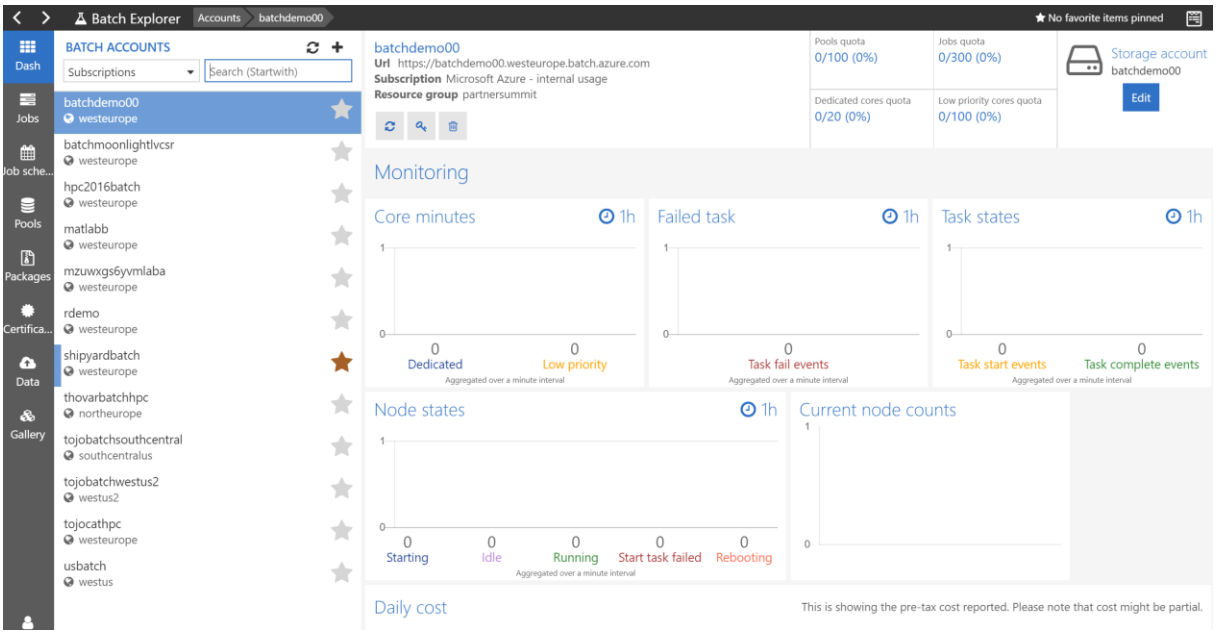


5. Sign in using the account associated with your Azure Subscription.



6. Wait for Batch Explorer to complete loading. When the main Batch Explorer window opens, in the left panel you should be able to select your Azure subscription and the just created Batch

account.



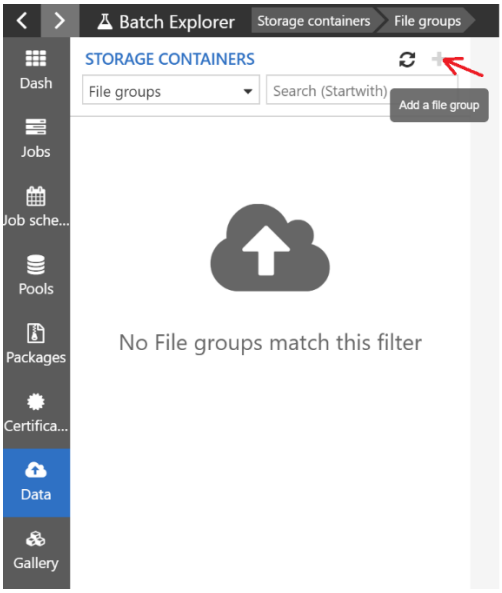
7. Your instance Batch Explorer is ready for use.

Now we will use the just installed Batch Explorer to process in parallel a set of input jpeg files with ImageMagick application.

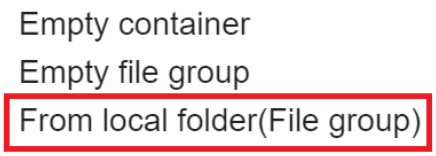
Step 1: Upload input files to Batch Labs input file group

1. Download sample set of input jpeg files from https://github.com/tojozefi/scfelab/blob/master/pictures.zip
2. Uncompress the content to any local folder, preserving the internal directory structure – folder 'Pictures' with 8 jpeg files.

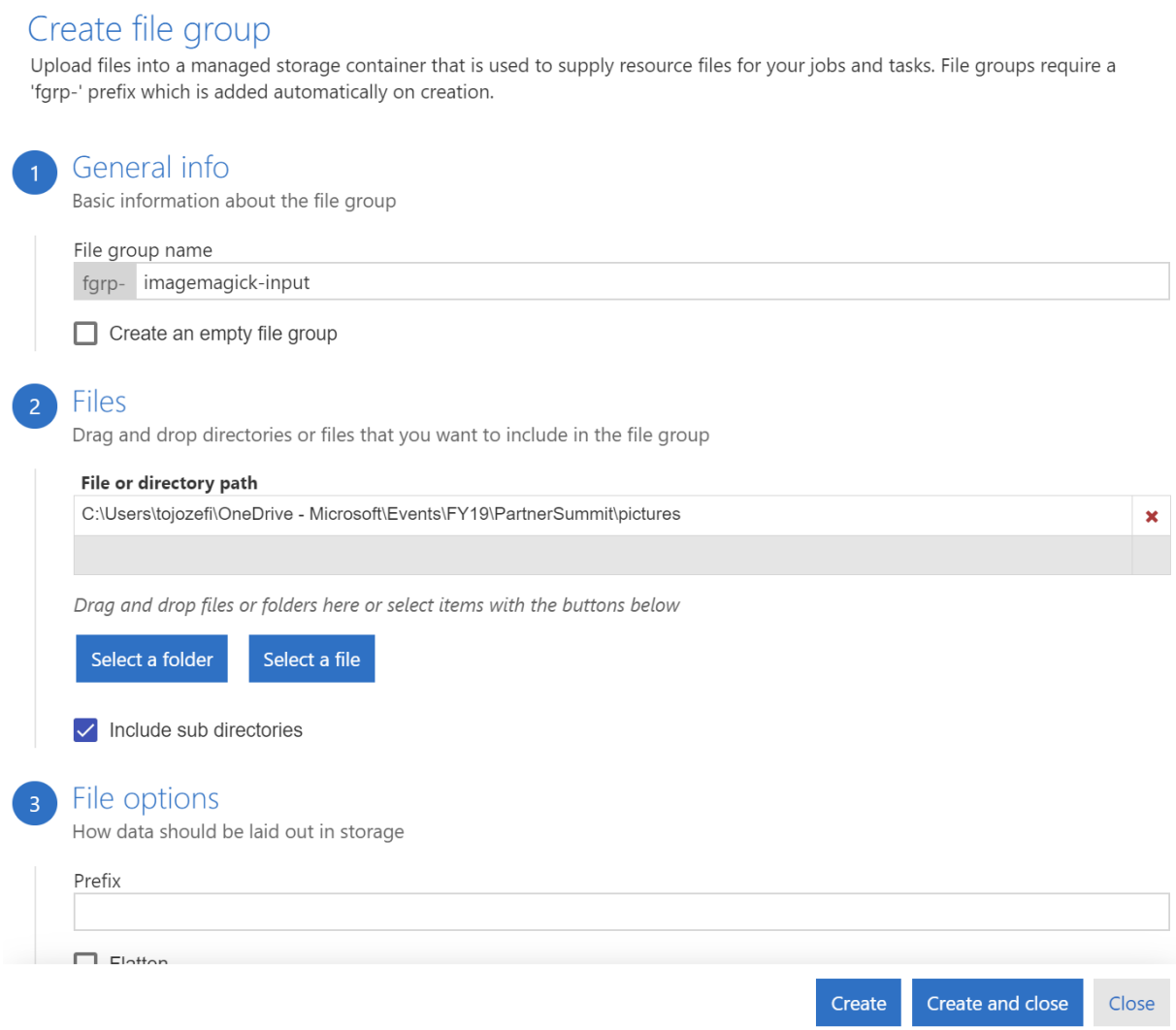
3. Navigate to **Data** panel in Batch Explorer and click **Add file group (+)** to create a new storage container:



4. Select **From local folder (file group)** from the drop-down list:

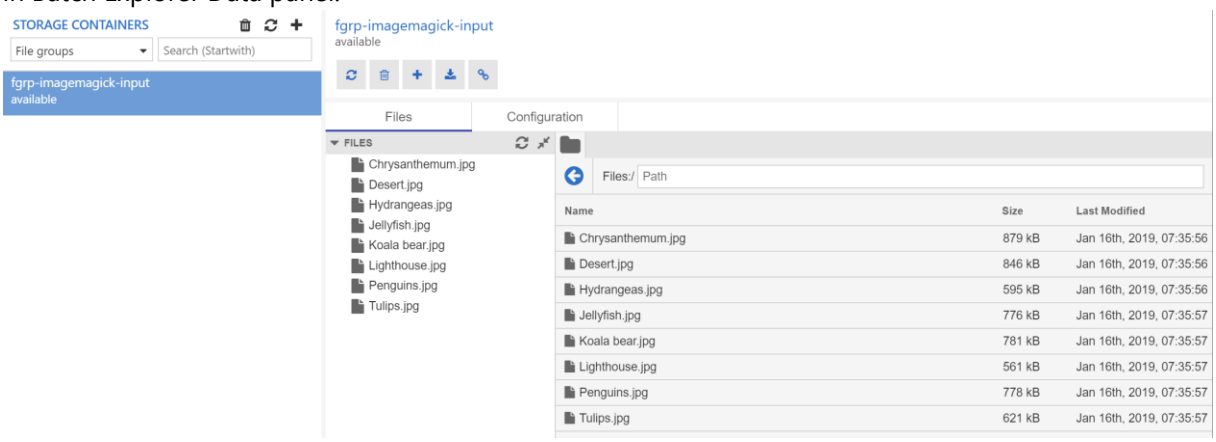


5. Fill out the **Create file group** form that opens:



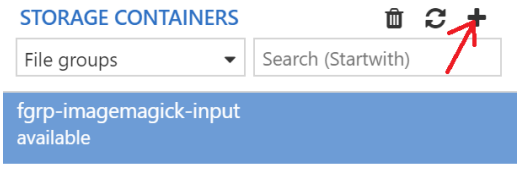
In (1**) General info** provide the name for the input file group, e.g. **imagemagick-input**
In (2) **Files** select the *pictures* directory with sample jpeg files from local disk and check **Include sub directories** box.
Leave all options in (3) **File options** intact.

6. Click **Create and close** button and wait for the file upload to complete.
On successful upload you should see all jpeg files in the **fgrp-imagemagick-input** file group in Batch Explorer Data panel:

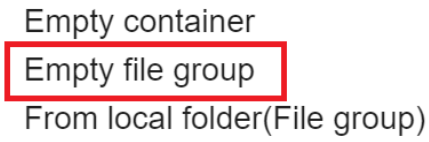


Step 2: Create output file group

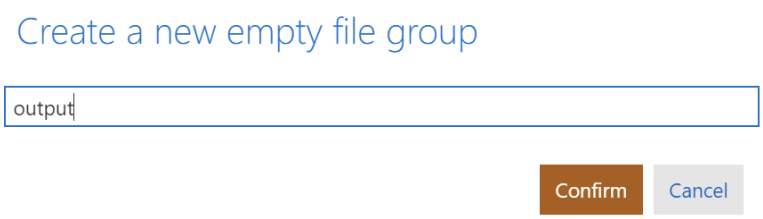
7. Click again **Add file group (+)** to create an output storage container:



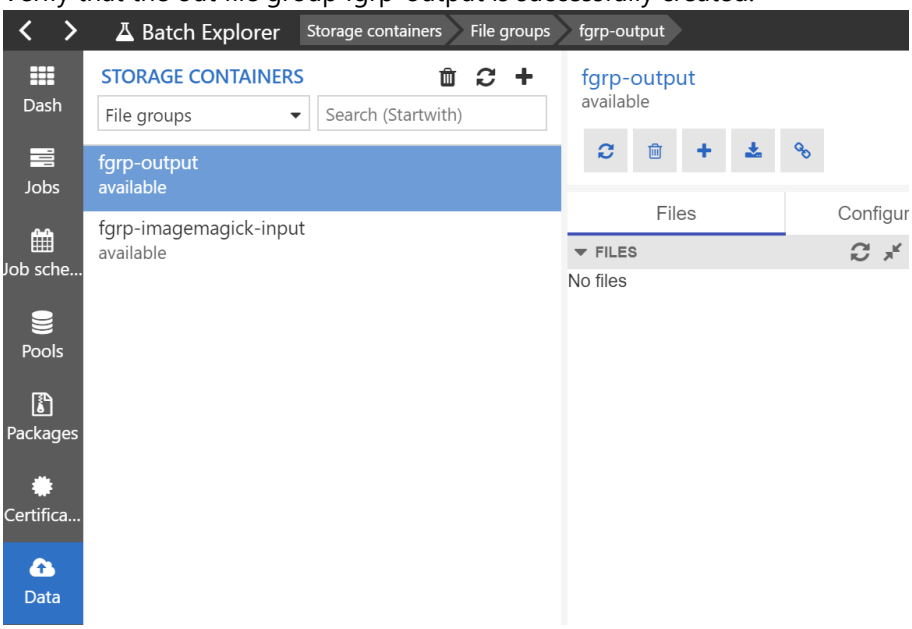
8. Select **Empty file group** from the drop-down list:



9. Provide a name for the output file group, e.g. **image-magick-output** and confirm.

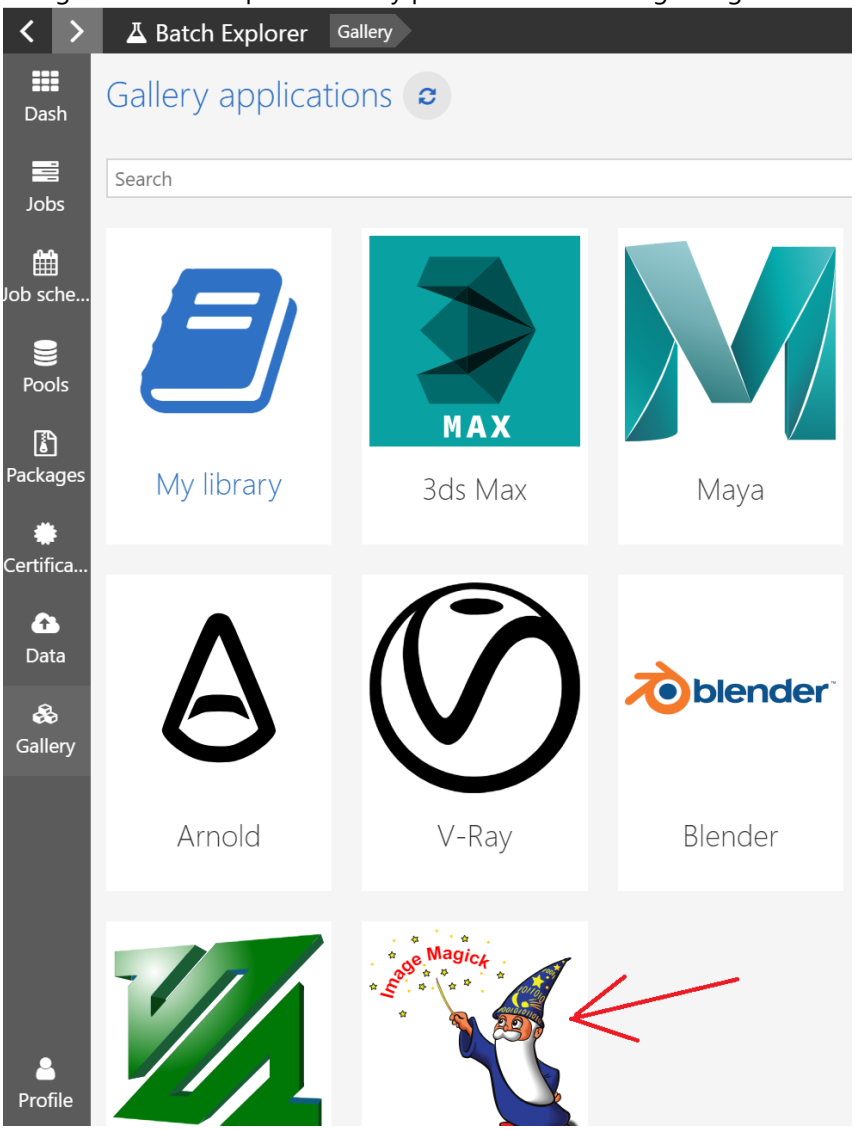


10. Verify that the out file group fgrp-output is successfully created:

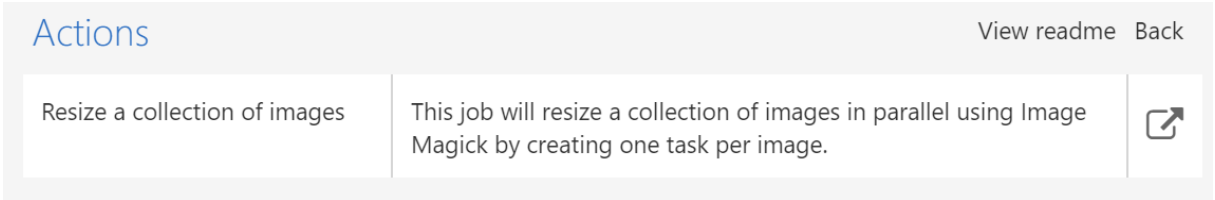


Step 3: Start Image Magick resize job for provided input files

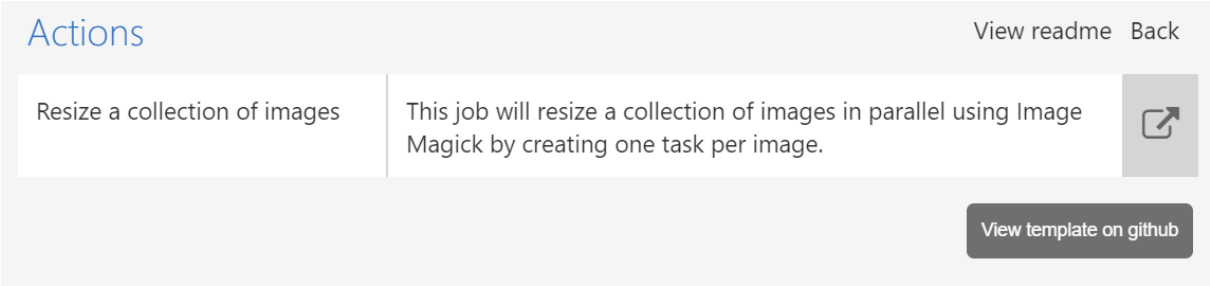
11. Navigate to Batch Explorer Gallery panel and select Image Magick tab:



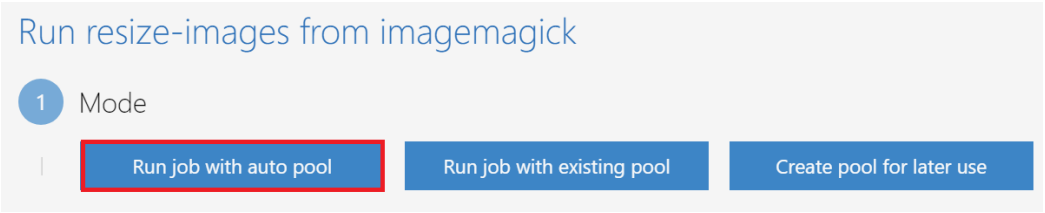
12. Select action **Resize a collection of images**:



Note: you can click on **view readme** to visit a website with ImageMagick Batch template repository.

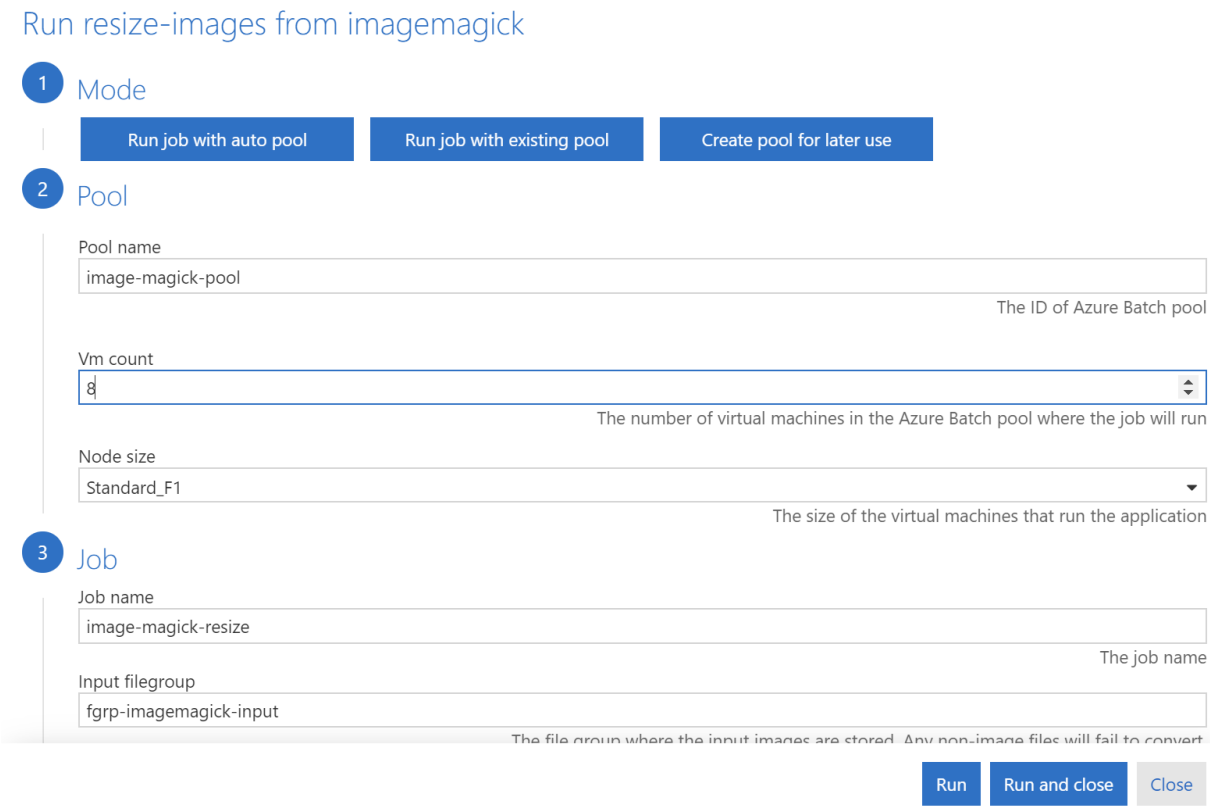


13. Select **Run job with auto pool** in (1) Mode selection:



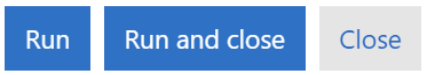
The job will be executed in automatically managed Azure Batch pool – i.e. the pool will be automatically deployed before the job start and deleted after the job is finished.

14. Fill out the remaining configuration parameters:

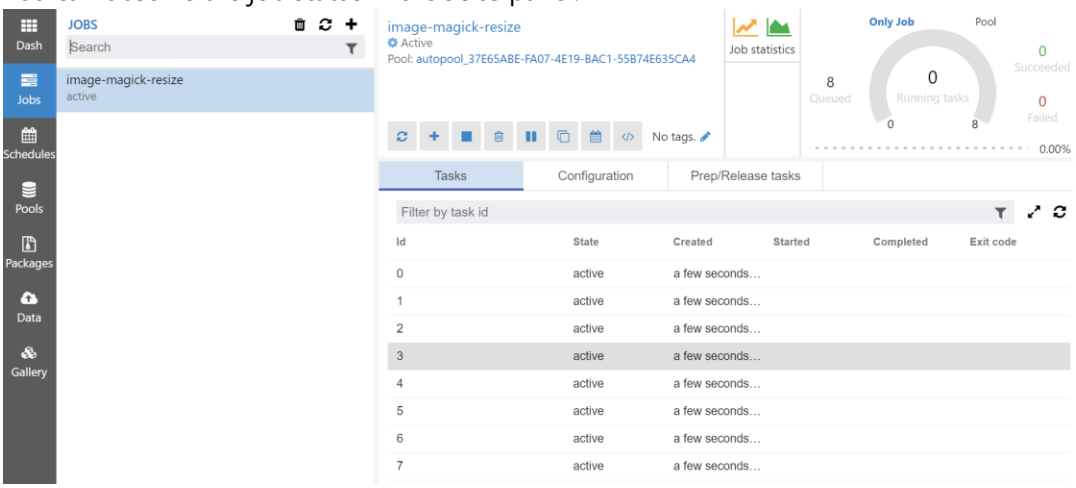


In (2) **Pool** definition provide a Batch pool name, e.g. **image-magick-pool**, **VM count** equal to the number of input files (in our case **8**) and select **VM size STANDARD_F1** from the list. In (3) **Job** definition provide the job name, e.g. **image-magick-resize**, and select the existing input and output file groups from the drop-down list – **fgrp-imagemagick-input** and **fgrp-image-magick-output** respectively. You may leave the **Resize** factor of **50%** or adjust.

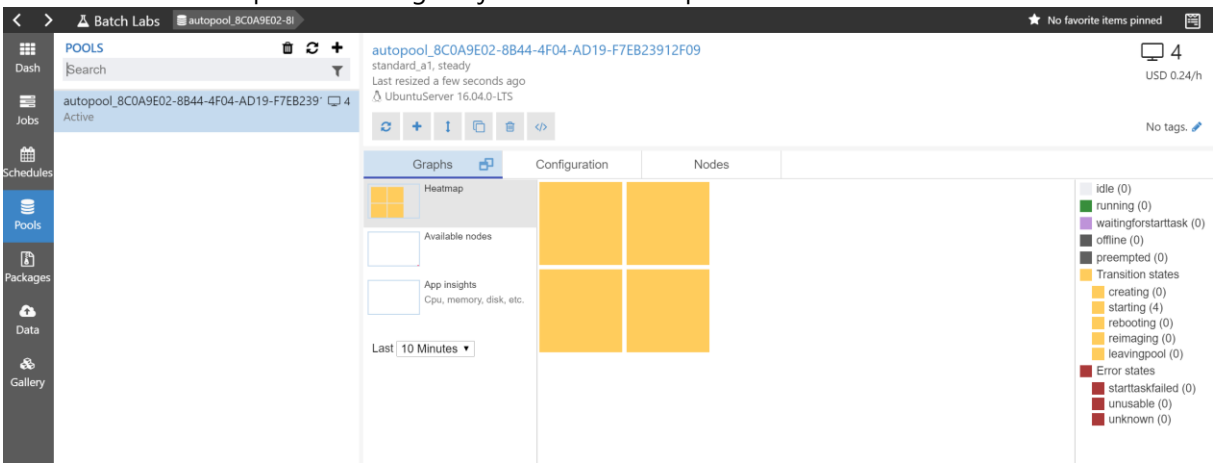
15. When configuration is ready, click **Run and close** button to start the job:



16. Monitor the job until it's finished.
    You can observe the job status in the **Jobs** panel:



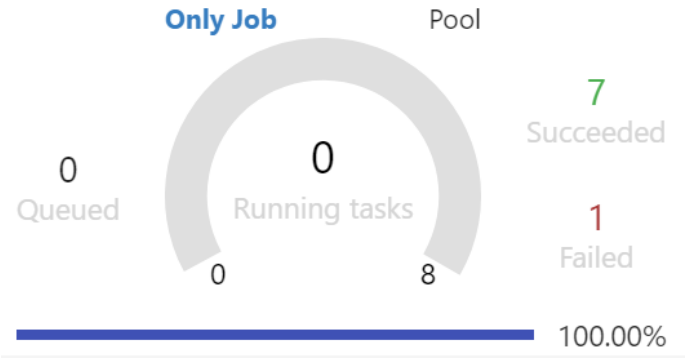
    and the status of the pool executing the job on the **Pools** panel:



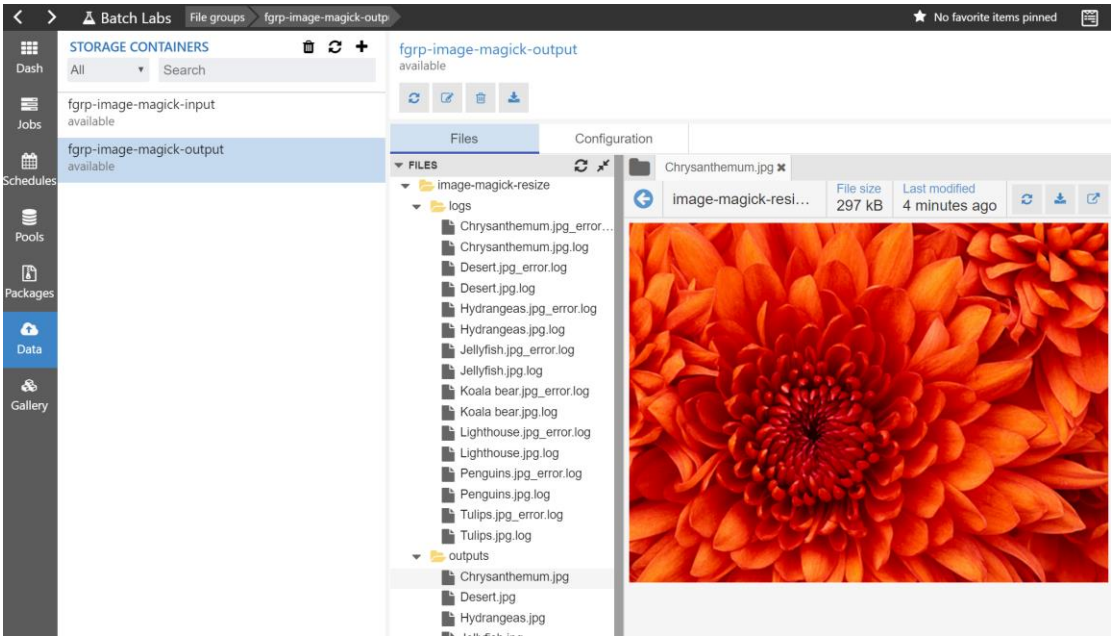
    The job should take about 5 minutes to complete.

Step 4: Check result of Image Magick resize job

17. When the job is finished you should see the following status on the Jobs panel (top-right corner):

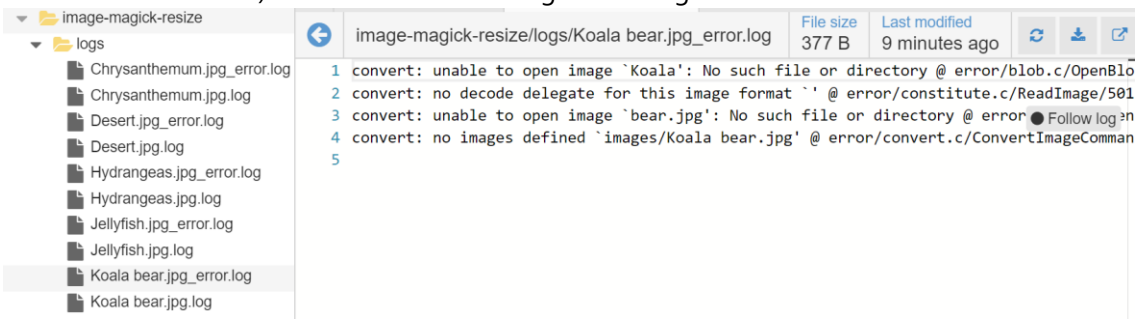


18. Navigate to Data panel in Batch Explorer main window and open fgrp-output file group. You should find image-magick-resize job's folder with log and output subfolders:

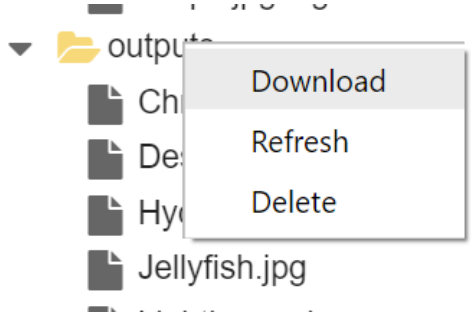


Check that the output jpeg files have reduced sizes as expected.

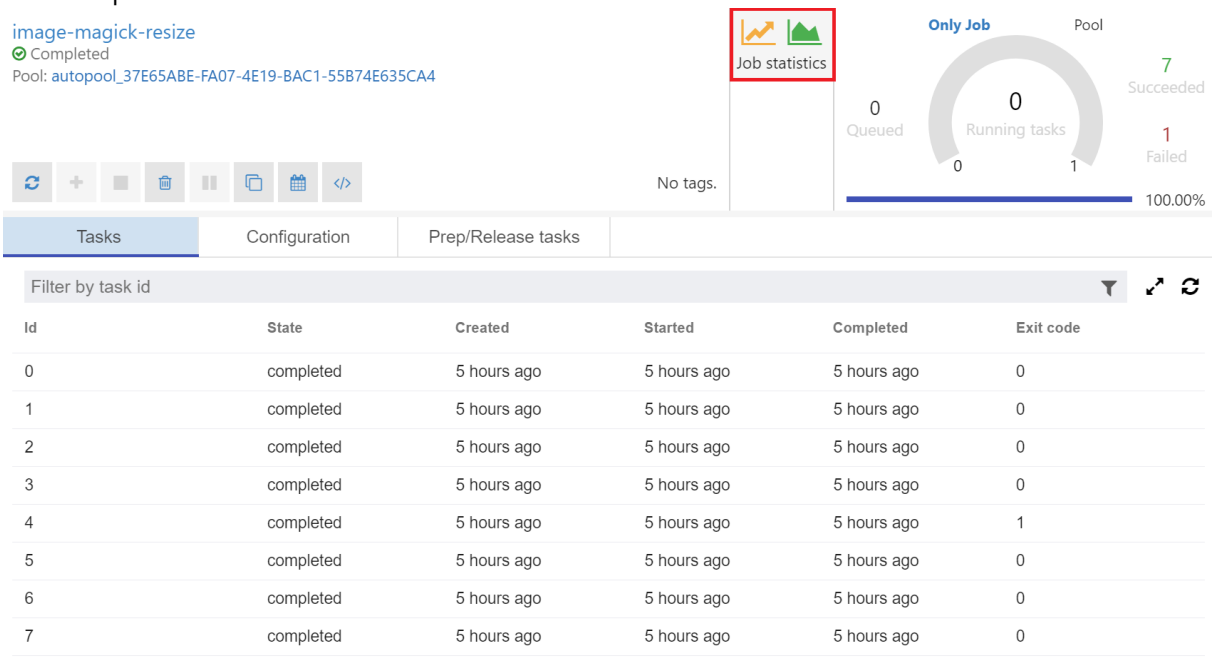
19. You may have observed that one of the tasks failed (for file Koala bear.jpg – due to space character in the name). Check the error message in the log file:



20. You can download the output files to local disk by right-clicking the **output** folder and selecting download action:

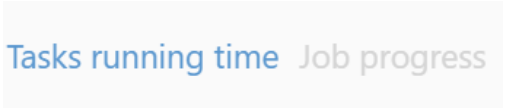


21. You can check job execution statistics by clicking **Job statistics** for *image-magick-resize* job on the **Jobs** panel:

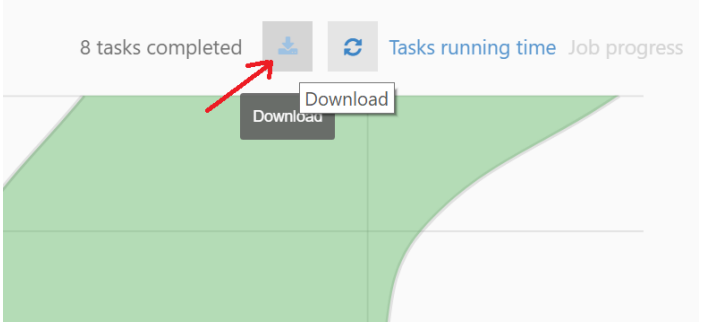


Id	State	Created	Started	Completed	Exit code
0	completed	5 hours ago	5 hours ago	5 hours ago	0
1	completed	5 hours ago	5 hours ago	5 hours ago	0
2	completed	5 hours ago	5 hours ago	5 hours ago	0
3	completed	5 hours ago	5 hours ago	5 hours ago	0
4	completed	5 hours ago	5 hours ago	5 hours ago	1
5	completed	5 hours ago	5 hours ago	5 hours ago	0
6	completed	5 hours ago	5 hours ago	5 hours ago	0
7	completed	5 hours ago	5 hours ago	5 hours ago	0

You can view task running times and job progress diagrams:

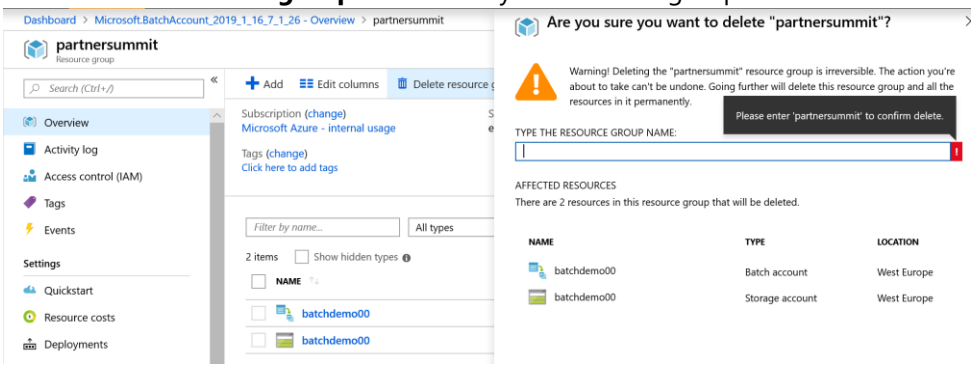


You can also download the job execution data to csv file by clicking download icon:



Step 5: Clean up resources

22. As the last step of the exercise clean up the created resources by deleting the resource group. Click "**Delete resource group"** button on your resource group tab:



23. Enter the resource group name to confirm resource deletion and click **Delete**: