

01

Overview of Homework 1b

HW 1b – Extract & Load





Deploy a template

(Every Homework)



01

Open Azure
CloudShell



02

Clone Github
Repository



03

Create Azure
Template



04

Deploy
Azure
Template

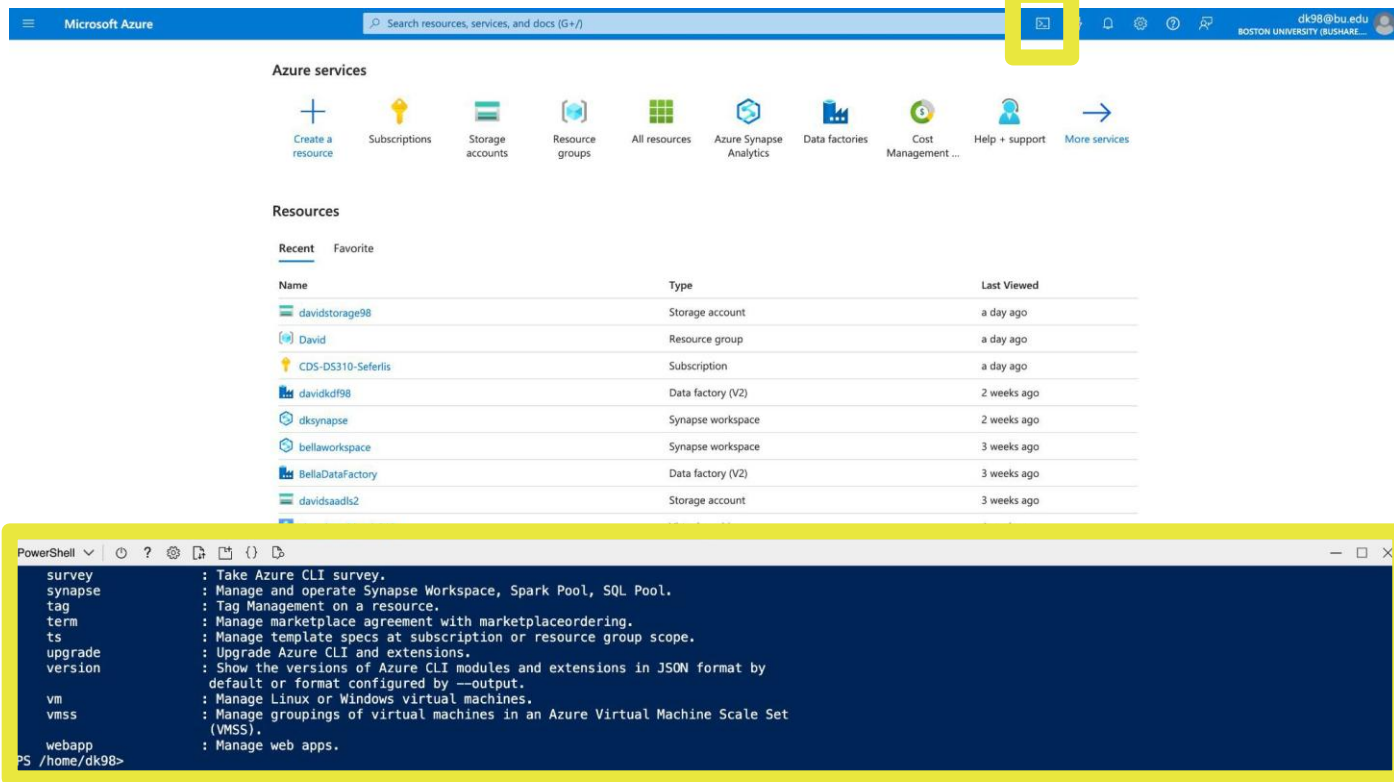


Why?



\$100 Free Credit is not a lot, and your **custom resource settings** may drain it even faster!

1. Open AZ Cloud shell



Microsoft Azure

Search resources, services, and docs (G+)

Cloud Shell

Azure services

- Create a resource
- Subscriptions
- Storage accounts
- Resource groups
- All resources
- Azure Synapse Analytics
- Data factories
- Cost Management
- Help + support
- More services

Resources

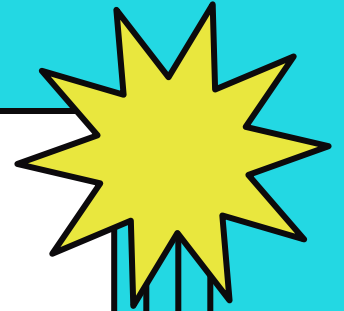
Recent Favorite

Name	Type	Last Viewed
davidstorage98	Storage account	a day ago
David	Resource group	a day ago
CDS-DS310-Seferlis	Subscription	a day ago
davidkf98	Data factory (V2)	2 weeks ago
dksynapse	Synapse workspace	2 weeks ago
bellaworkspace	Synapse workspace	3 weeks ago
BellaDataFactory	Data factory (V2)	3 weeks ago
davidsaads2	Storage account	3 weeks ago

```
PowerShell
survey : Take Azure CLI survey.
synapse : Manage and operate Synapse Workspace, Spark Pool, SQL Pool.
tag : Tag Management on a resource.
term : Manage marketplace agreement with marketplaceordering.
ts : Manage template specs at subscription or resource group scope.
upgrade : Upgrade Azure CLI and extensions.
version : Show the versions of Azure CLI modules and extensions in JSON format by default or format configured by --output.
vm : Manage Linux or Windows virtual machines.
vmss : Manage groupings of virtual machines in an Azure Virtual Machine Scale Set (VMSS).
webapp : Manage web apps.
PS /home/dk98>
```

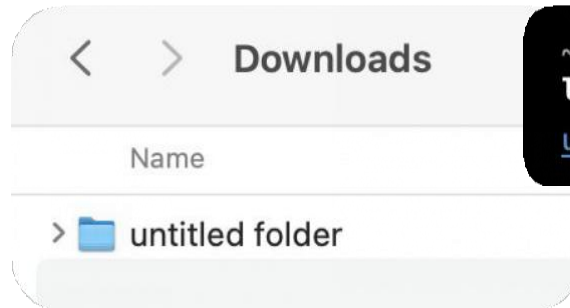
Azure Cloud shell

- **Magic Shell**
 - Can **ask the sea** to do things like bring you a wave, calm down, or even change the tide.
- **Cloud Shell**
 - Can **ask the cloud** to start services, manage them, or even troubleshoot issues, just by typing in commands.

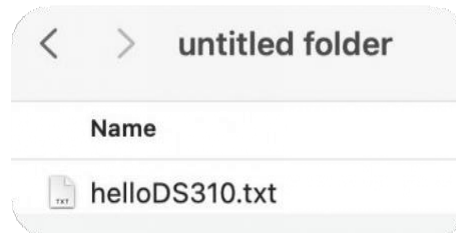


Linux Command

- **cd (Change Directory)**
 - change the current working directory
- **ls (List)**
 - list the contents of a directory

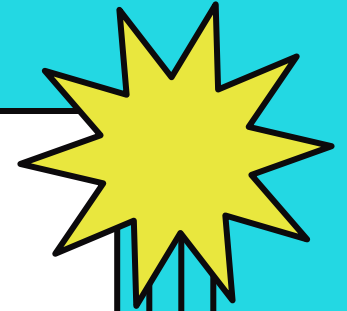


```
~/Downloads (0.028s)  
ls  
untitled folder
```

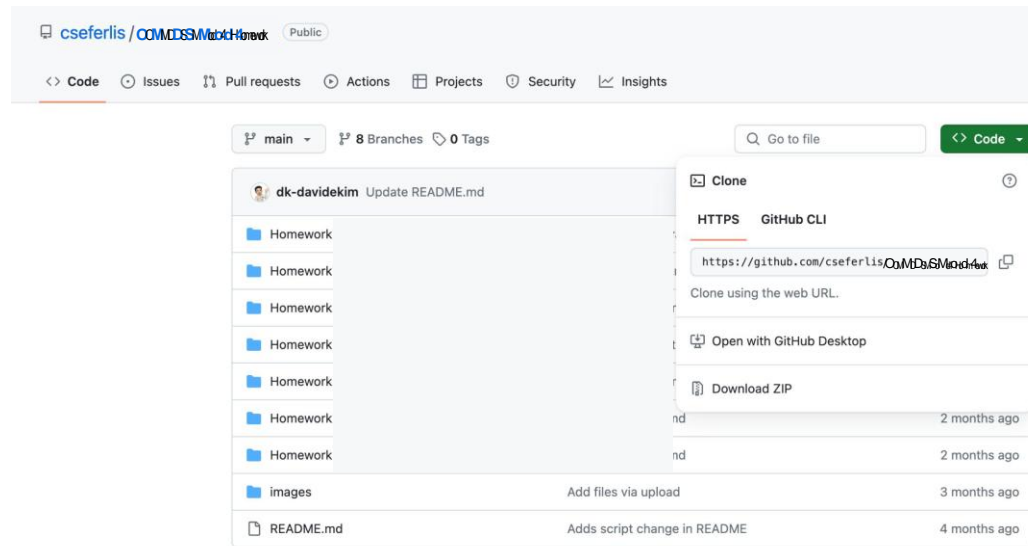


```
~/Downloads (0.022s)  
cd 'untitled folder'  
~/Downloads/untitled folder
```

```
~/Downloads/untitled folder  
ls  
helloDS310.txt
```



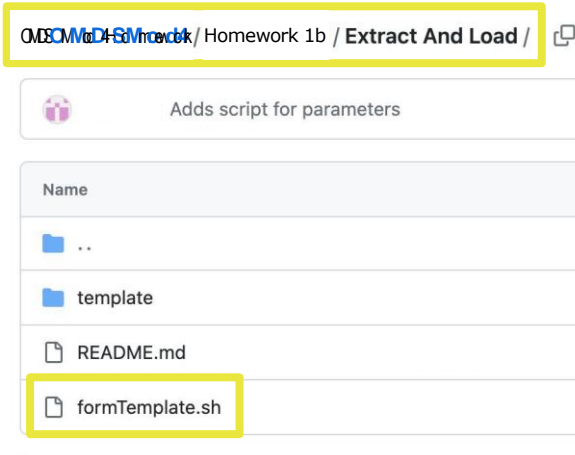
2. Clone Github repo



<https://github.com/cseferlis/COMDSMod4Homework.git>

```
PS /home/dk98> git clone https://github.com/cseferlis/ds310homework.git
```


3. Navigate to formTemplate.sh



PS /home/seferlis/OMDSMod4Homework / Homework 1b / Extract And Load >

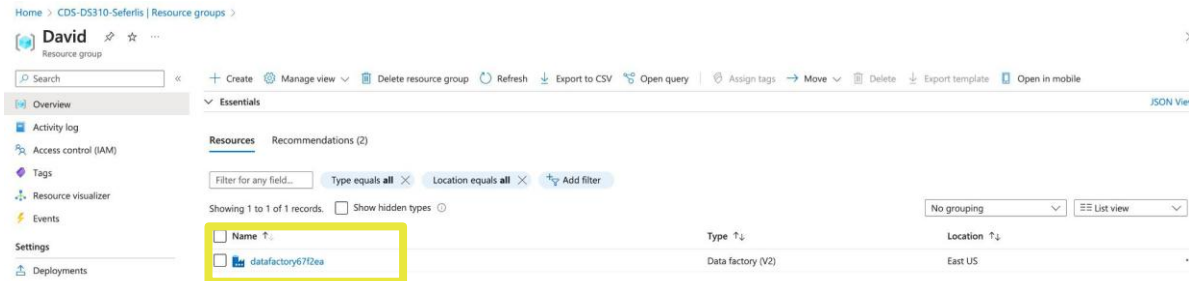
bash ./formTemplate.sh

Template and parameters created successfully: ./template/template.json and ./template/parameters.json

4. Deploy Azure Template

Template and parameters created successfully: ./template/template.json and ./template/parameters.json

- `az deployment group create --resource-group ResourceGroupName --template-file TemplateLocation --parameters ParametersLocation`
 - e.g. Resource Group: David
 - `az deployment group create --resource-group David --template-file ./template/template.json --parameters ./template/parameters.json`





Homework 1b

Extract & Load



01

Data Factory



02

Extract & Load

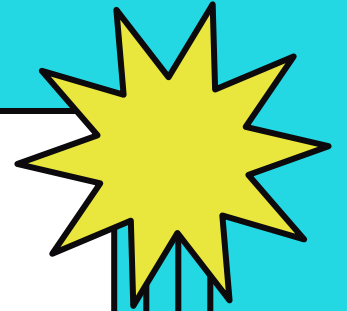


03

Unzip

1. Data Factory

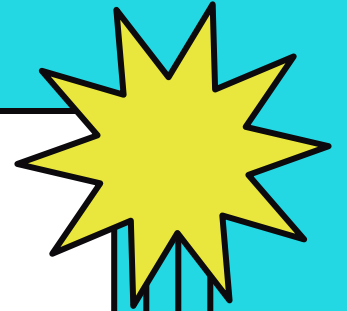
- **Large kitchen** with recipes coming from all over the world
 - You're one of the **Chefs**
 - **Gathering Ingredients** (Data Ingestion)
 - **Preparing Ingredients** (Data Transformation)
 - **Cooking** (Data Computation)
 - **Serving the Dish** (Data Publication)



1. Data factory

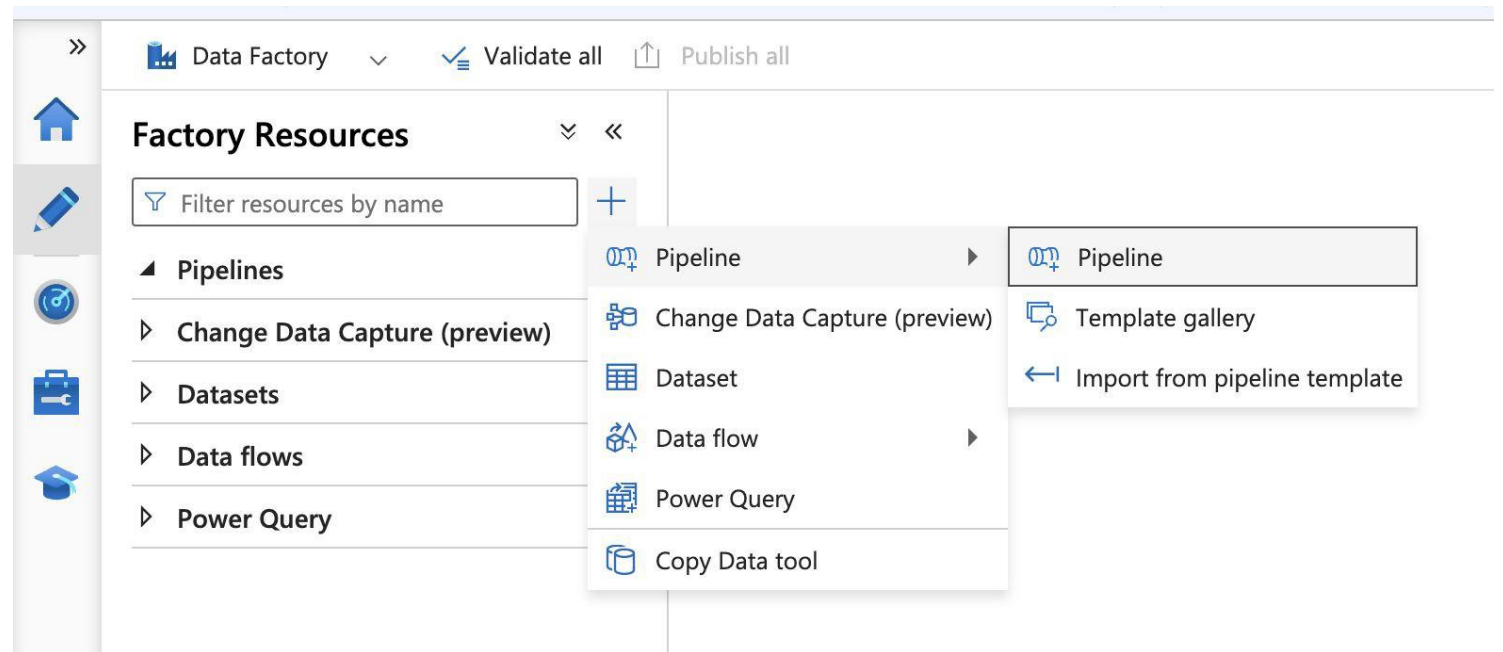
Watch this video at home!

Azure Data Factory Video



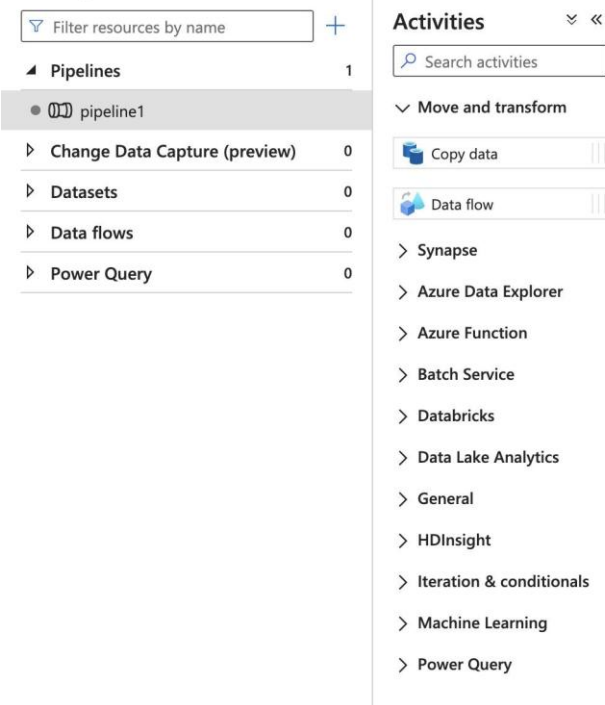
Homework 1b

Creating a Pipeline



Homework 1b

Creating Activities



01

COPY DATA 1

02

COPY DATA 2



Extract & Load

Unzip

2. Extract & Load

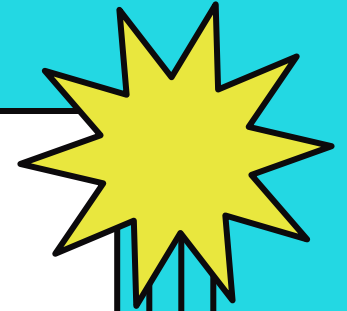
- Set up **Copy Data** activity to **Extract data** from
 - [‘https://static.nhtsa.gov/odi/ffdd/cmpl/FLAT_CMPL.zip’](https://static.nhtsa.gov/odi/ffdd/cmpl/FLAT_CMPL.zip)
 - Hint: Zip file is a **binary** file!
- 1. Extract it from **HTTP** Linked Service
- 2. Load it to a container in your **Azure Blob Storage**



HTTP



Azure Blob Storage



3. Unzip

- **Unzip the file** in the Azure Blob Storage
- Hint:
 - Change the compression type
 - Change the .zip extension to...
 - Check Azure Storage Account Container for Output
 - Size of the txt file should be larger than that of the zip file



Azure Blob Storage

*Running a pipeline

- **Publish & Add Trigger** >>Cheaper than 'Debug' Action
- Monitor pipeline runs through the '**monitor**' section
 - Go to details (eyeglasses icon) for details on data read & written)

The screenshot displays the 'Add trigger' dropdown menu on the left, which includes options for 'Trigger now' and 'New/Edit'. In the center is the navigation pane with 'Pipeline runs' selected. On the right, the 'Activity runs' section shows a single run with ID '21c23e6d-83e4-4684-aa20-a0e5562acef8' that has 'Succeeded'.

Add trigger

- Trigger now
- New/Edit

Navigation:

- Dashboards
- Runs
- Pipeline runs (selected)
- Trigger runs
- Change Data Capture (previ...

Activity runs

Pipeline run ID 21c23e6d-83e4-4684-aa20-a0e5562acef8

All status ▾

Showing 1 - 1 items

Activity name	Activity status
Copy_4p9	Succeeded

Details



Best of luck!