01 OVERVIEW OF HW 2

HW 2 - Extract & Load

DEPLOY A TEMPLATE

(Every Homework)

01

Open Azure CloudShell

02

Clone Github Repository

03

Create Azure Template

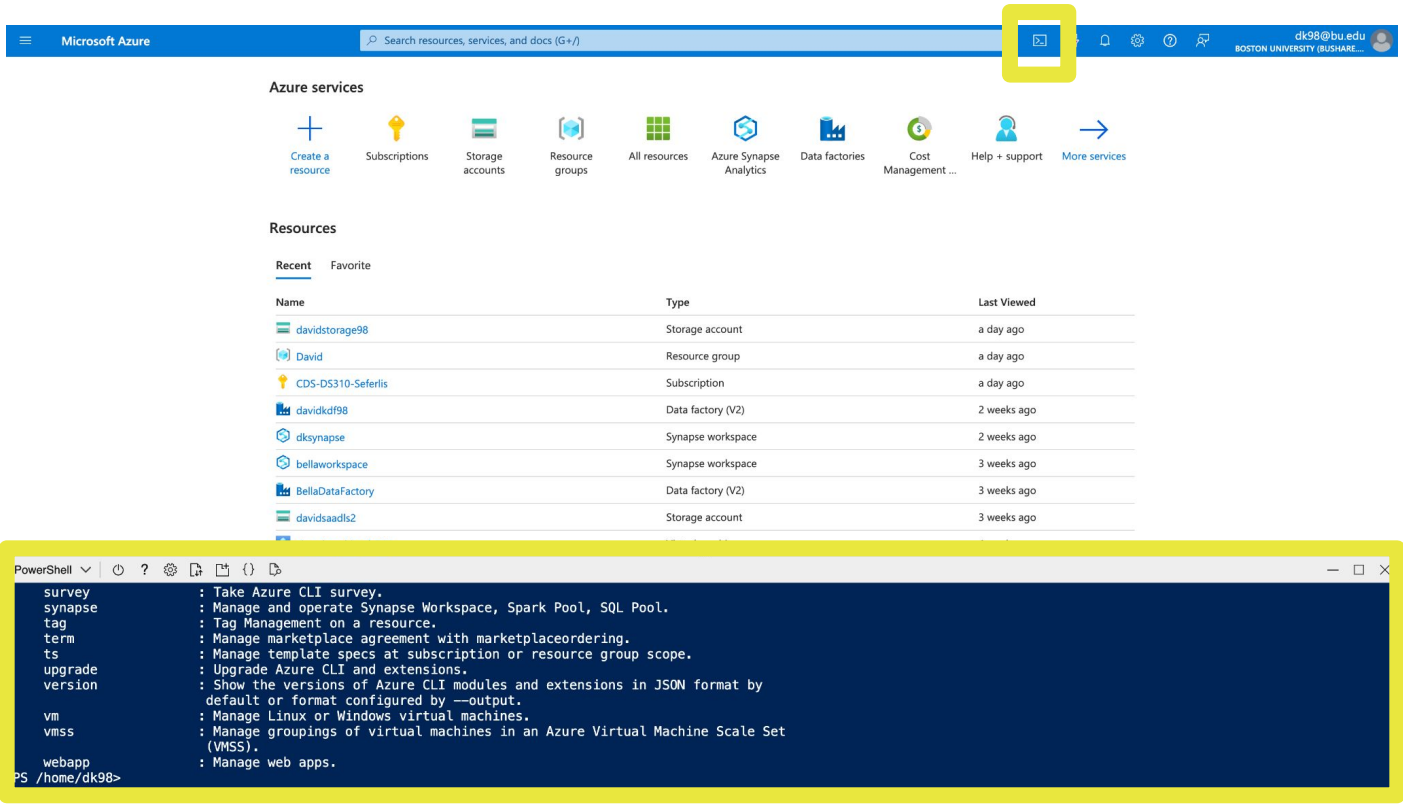
04

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



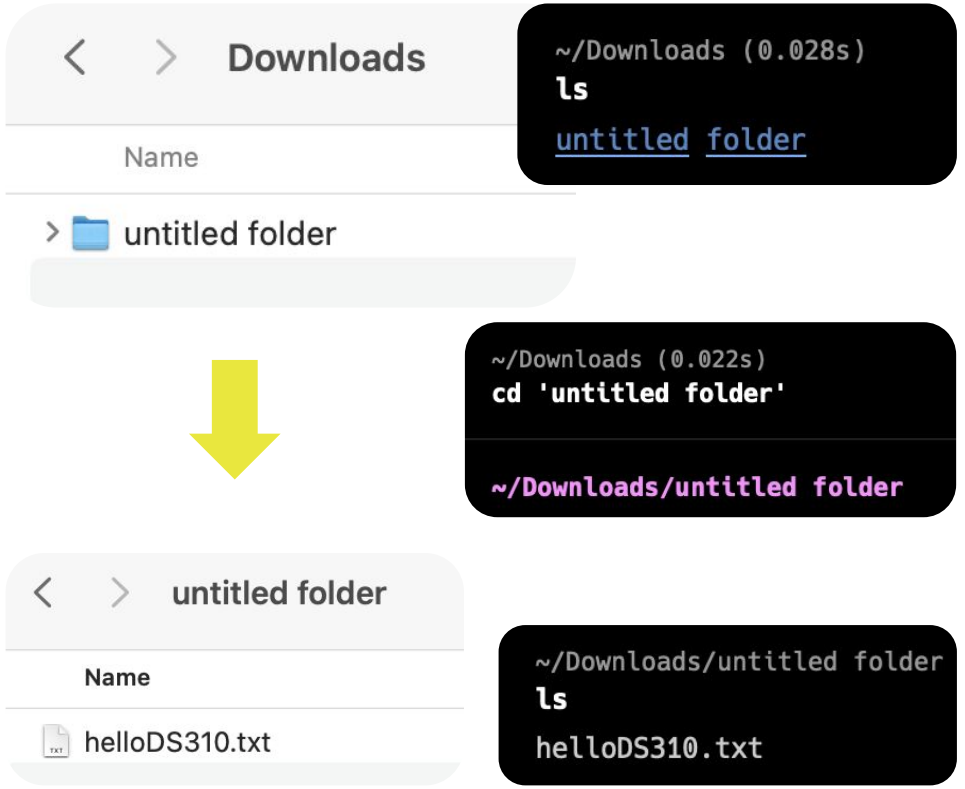
AZURE CLOUD SHELL

- **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.

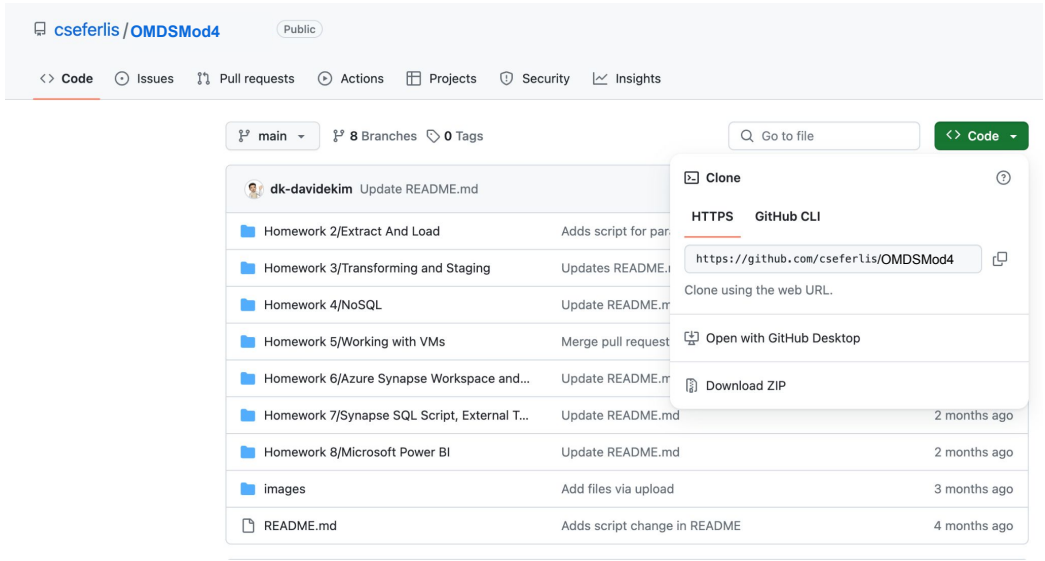
LINUX COMMAND

- **cd (Change Directory)**
  - change the current working directory

- **ls (List)**
  - list the contents of a directory



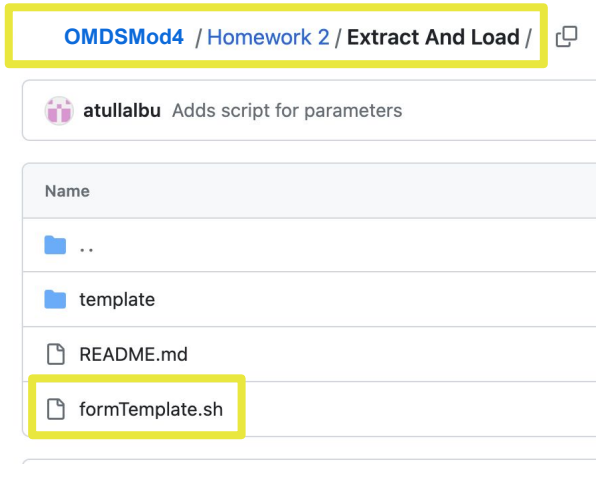
2. CLONE GITHUB REPO



https://github.com/cseferlis/OMDSMod4.git

```
PS /home/dk98> git clone https://github.com/cseferlis/ OMDSMod4 .git
```

3. NAVIGATE TO FORMTEMPLATE.SH




```
PS /home/dk98> cd OMDSMod4/'Homework 2'/'Extract And Load'
```

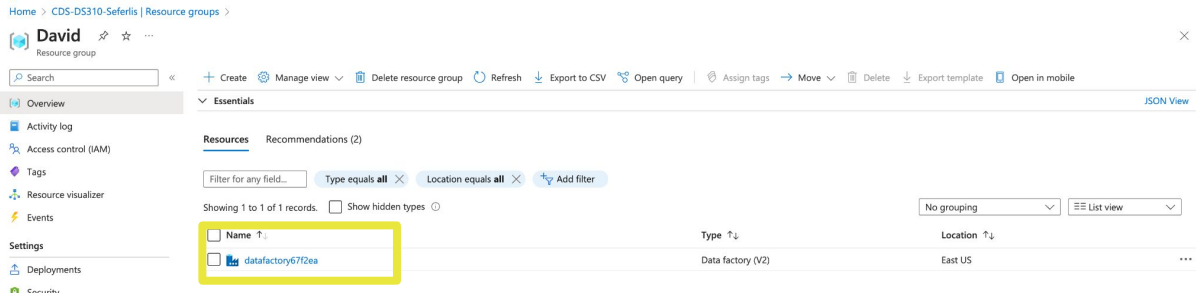
```
bash ./formTemplate.sh
```

```
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 2

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)

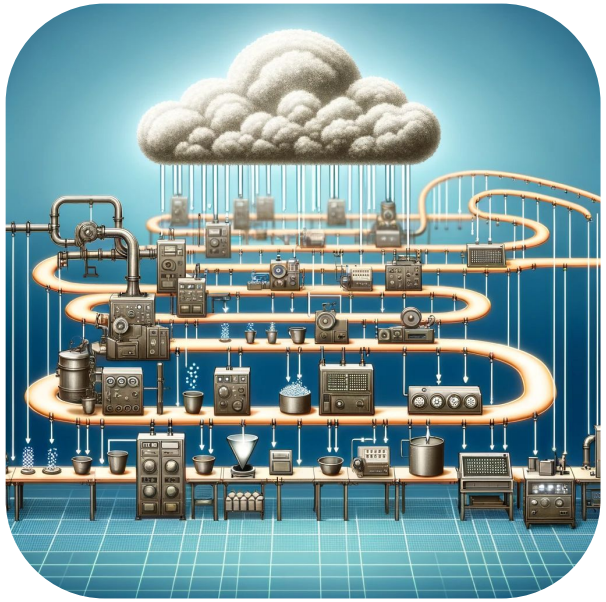
1. DATA FACTORY

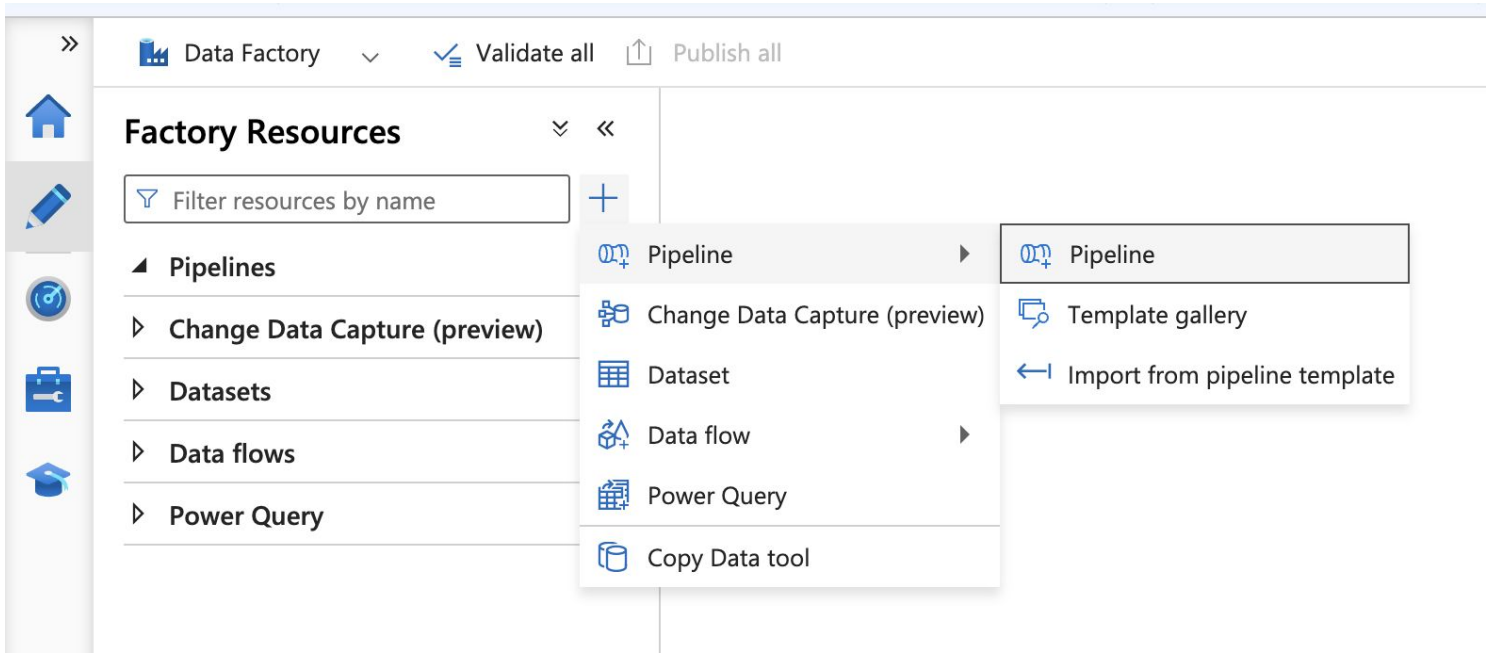
Watch this video at home!

https://www.microsoft.com/en-us/videoplayer/embed/RE4Mc3u

HOMEWORK 2

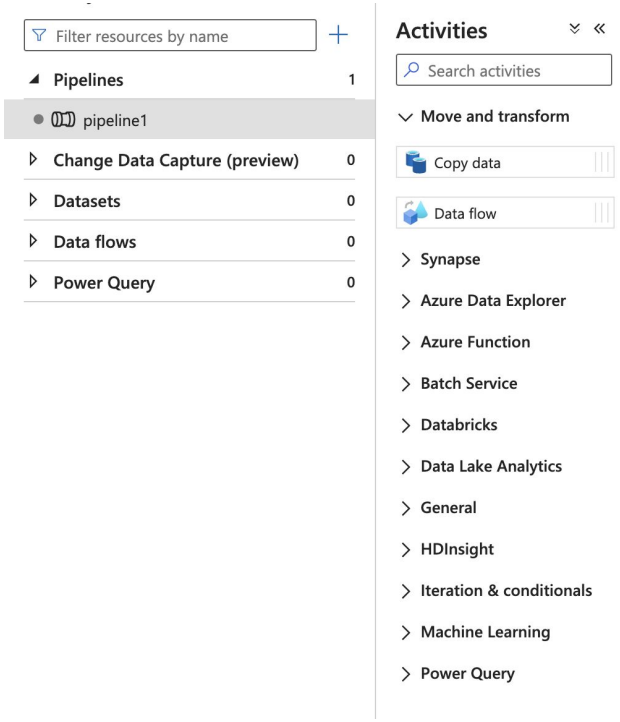
Creating a Pipeline





HOMEWORK 2

Creating Activities

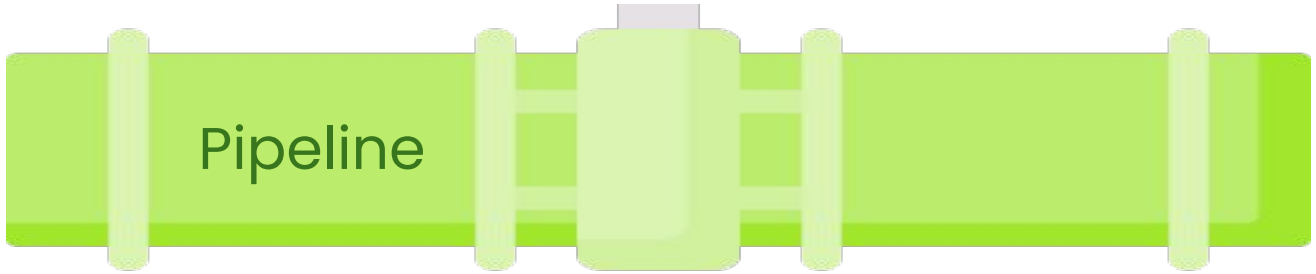


01

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’
  - Hint: Zip file is a **binary** file!

HTTP

Azure Blob Storage

1. Extract it from **HTTP** Linked Service
2. Load it to a container in your **Azure Blob Storage**

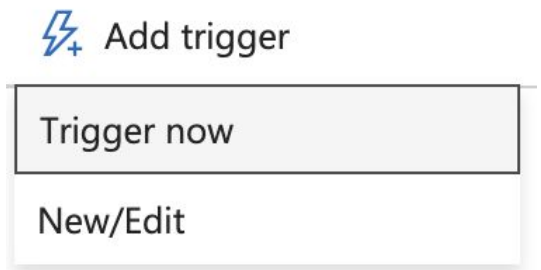
3. UNZIP

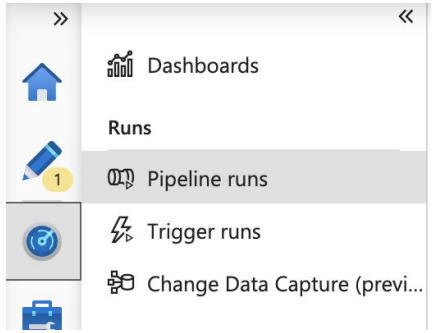
- **Unzip the file** in the Azure Blob Storage
- Hint:
  - Change the compression type
  - Change the .zip extension to...
  - Check Azure Storage Account Container for Output
    - Size of the txt file should be larger than that of the zip file

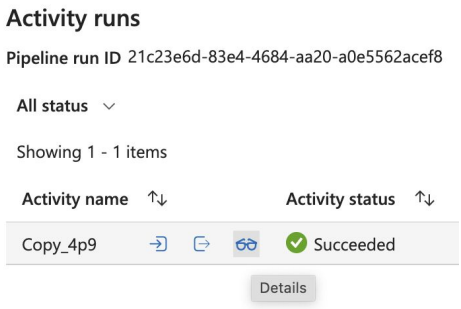
Azure Blob Storage

*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)







BEST OF LUCK!