



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

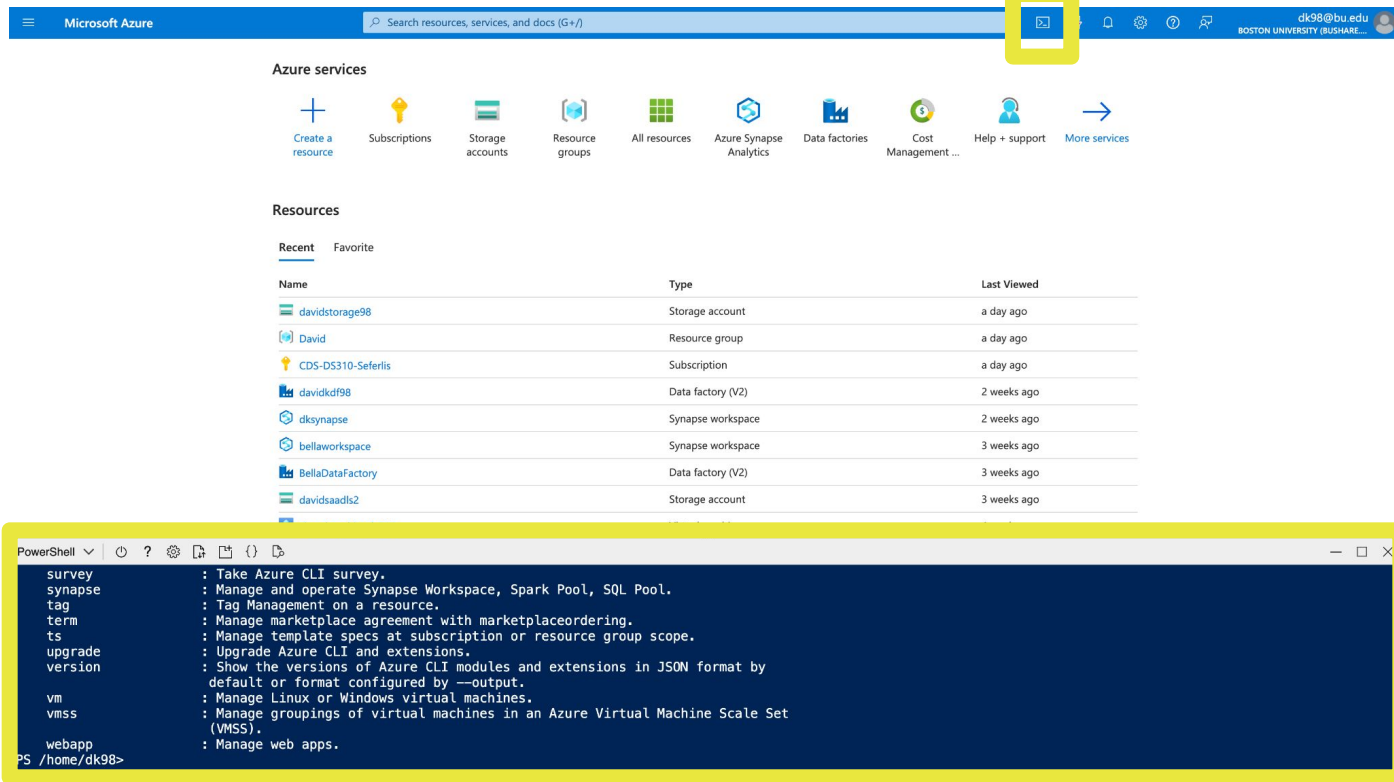


WHY?



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

1. OPEN AZ CLOUD SHELL



The screenshot shows the Microsoft Azure portal interface. In the top navigation bar, the 'Cloud Shell' icon (a terminal window) is highlighted with a yellow square. Below the navigation bar, the 'Azure services' section displays various icons for different services. The 'Resources' section shows a table of recent resources.

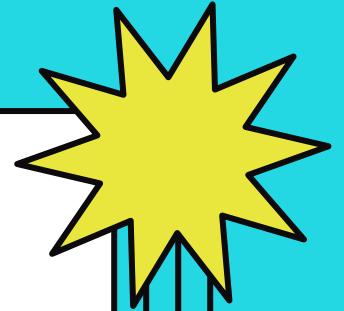
Name	Type	Last Viewed
davidstorage98	Storage account	a day ago
David	Resource group	a day ago
CDS-DS310-Seferlis	Subscription	a day ago
davidkdf98	Data factory (V2)	2 weeks ago
dksynapse	Synapse workspace	2 weeks ago
bellaworkspace	Synapse workspace	3 weeks ago
BellaDataFactory	Data factory (V2)	3 weeks ago
davidsaads2	Storage account	3 weeks ago

Below the resources table, a PowerShell terminal window is open, displaying the following commands and their descriptions:

```
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

The diagram illustrates the process of navigating into a directory and listing its contents using Linux commands. It consists of three terminal window snippets connected by a large yellow arrow pointing downwards.

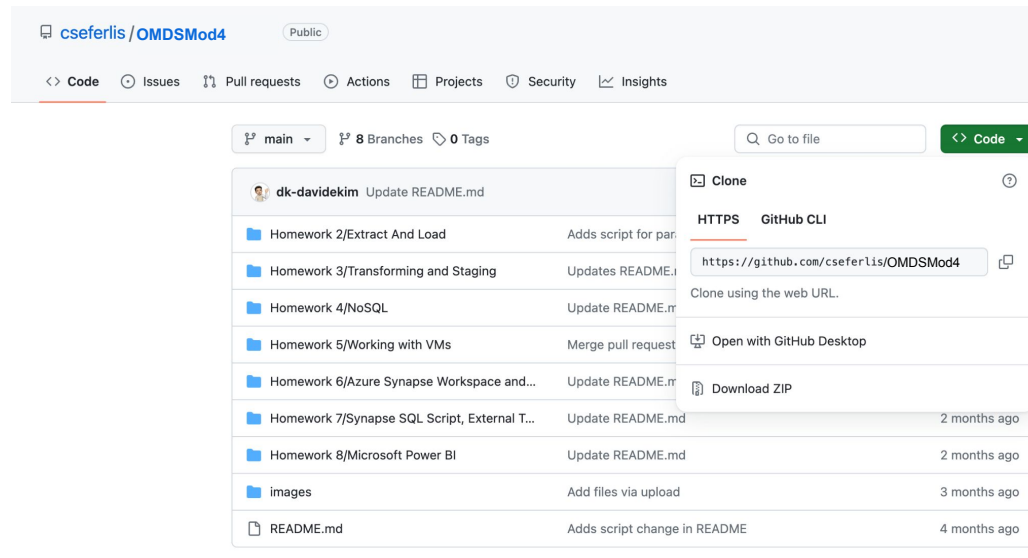
Terminal 1: Shows the initial directory as `Downloads`. The command `ls` is executed, listing `untitled folder`. The prompt is `~/Downloads (0.028s)`.

Terminal 2: Shows the command `cd 'untitled folder'` being executed. The prompt is `~/Downloads (0.022s)`. The current directory is now `~/Downloads/untitled folder`.

Terminal 3: Shows the command `ls` being executed in the new directory. The prompt is `~/Downloads/untitled folder`, and the output lists `helloDS310.txt`.

The visual elements include a yellow starburst in the top right corner and a purple spray can in the bottom right corner. The background features a blue grid pattern on the left and a yellow zigzag pattern at the bottom.

2. CLONE GITHUB REPO

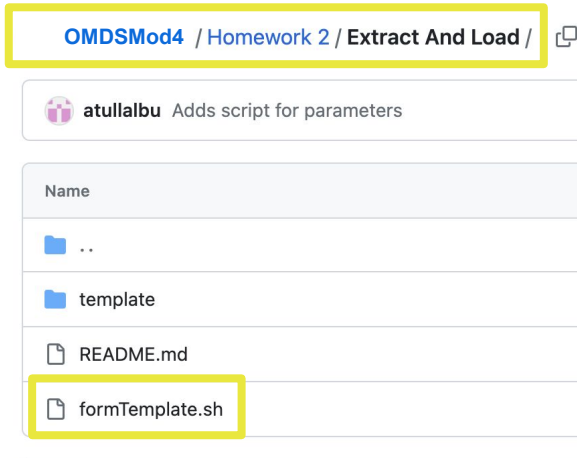


<https://github.com/cseferlis/OMDSMod4.git>



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


3. NAVIGATE TO FORMTEMPLATE.SH



PS /home/seferlis/OMDMod4/Homework 1b/Extract And Load>

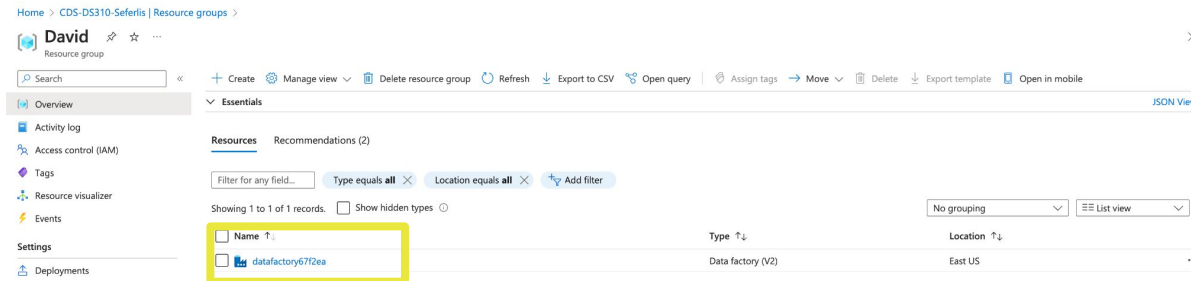
bash ./formTemplate.sh

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

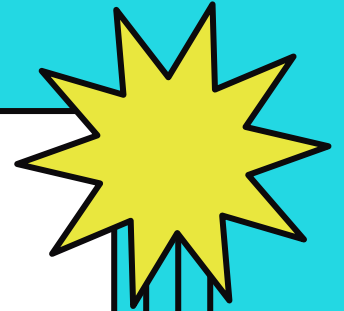


03

Unzip

1. DATA FACTORY

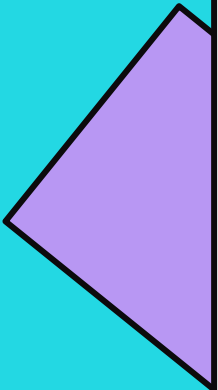
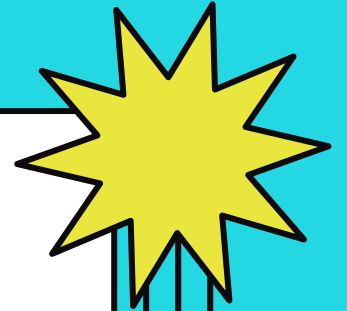
- **Large kitchen** with recipes coming from all over the world
 - You're one of the **Chefs**
 - **Gathering Ingredients** (Data Ingestion)
 - **Preparing Ingredients** (Data Transformation)
 - **Cooking** (Data Computation)
 - **Serving the Dish** (Data Publication)



1. DATA FACTORY

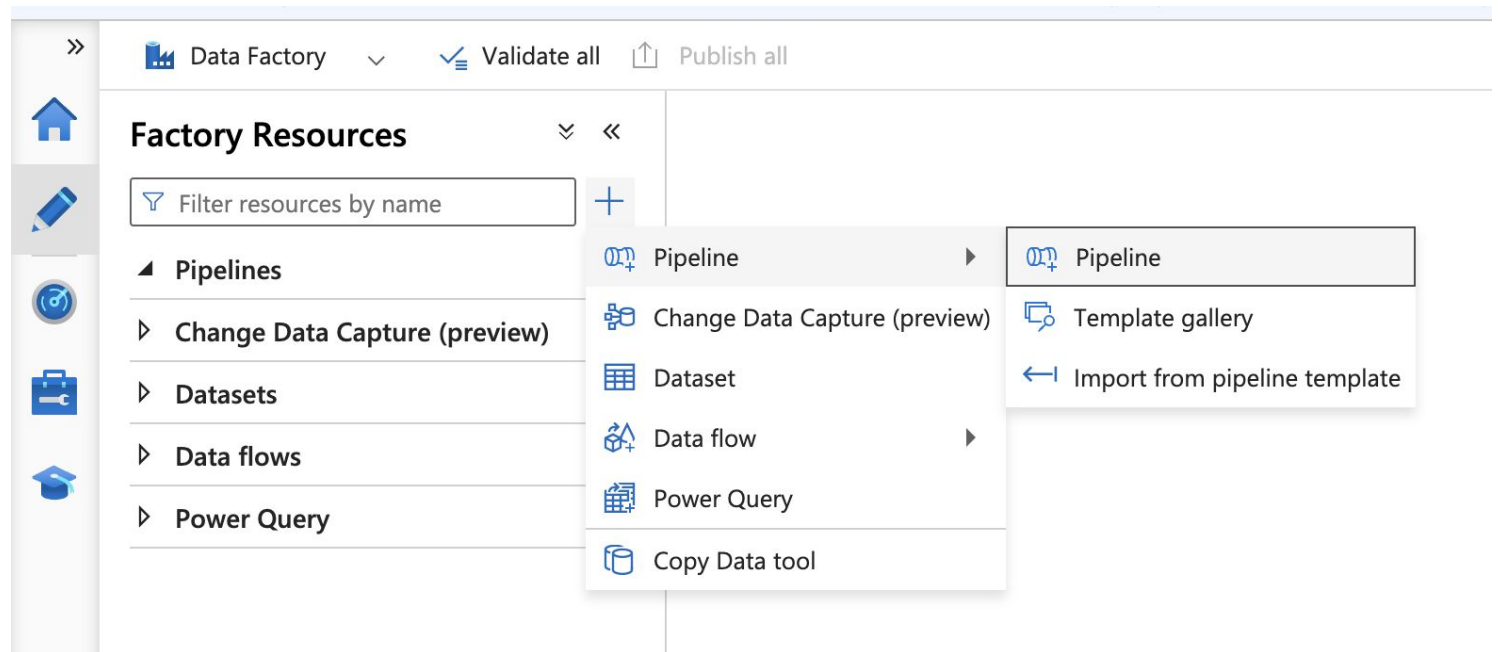
Watch this video at home!

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



HOMework 1B

Creating a Pipeline



HOMEWORK 1B

Creating Activities

+

Pipelines 1

● pipeline1

▸ Change Data Capture (preview) 0

▸ Datasets 0

▸ Data flows 0

▸ Power Query 0

Activities

Search activities

▼ Move and transform

Copy data

Data flow

> Synapse

> Azure Data Explorer

> Azure Function

> Batch Service

> Databricks

> Data Lake Analytics

> General

> HDInsight

> Iteration & conditionals

> Machine Learning

> Power Query

01

COPY DATA 1

02

COPY DATA 2



Extract & Load

Unzip

2. EXTRACT & LOAD

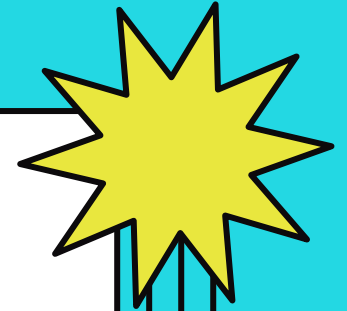
- Set up **Copy Data** activity to **Extract data** from
 - https://static.nhtsa.gov/odi/ffdd/cmpl/FLAT_CMPL.zip
 - Hint: Zip file is a **binary** file!
- 1. Extract it from **HTTP** Linked Service
- 2. Load it to a container in your **Azure Blob Storage**



HTTP



Azure Blob Storage

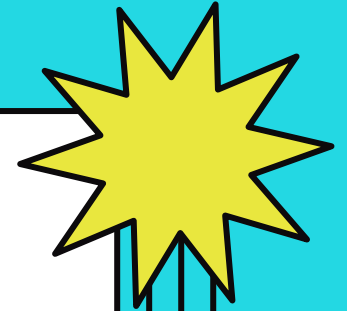


3. UNZIP

- **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 shows two parts of a software interface. On the left, the 'Add trigger' menu is open, showing options: 'Trigger now' and 'New/Edit'. On the right, the 'Activity runs' section is visible, showing a list of runs for a specific pipeline. The first run, 'Copy_4p9', is shown with a status of 'Succeeded' and a 'Details' button next to it.

Add trigger

- Trigger now
- New/Edit

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!

