



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

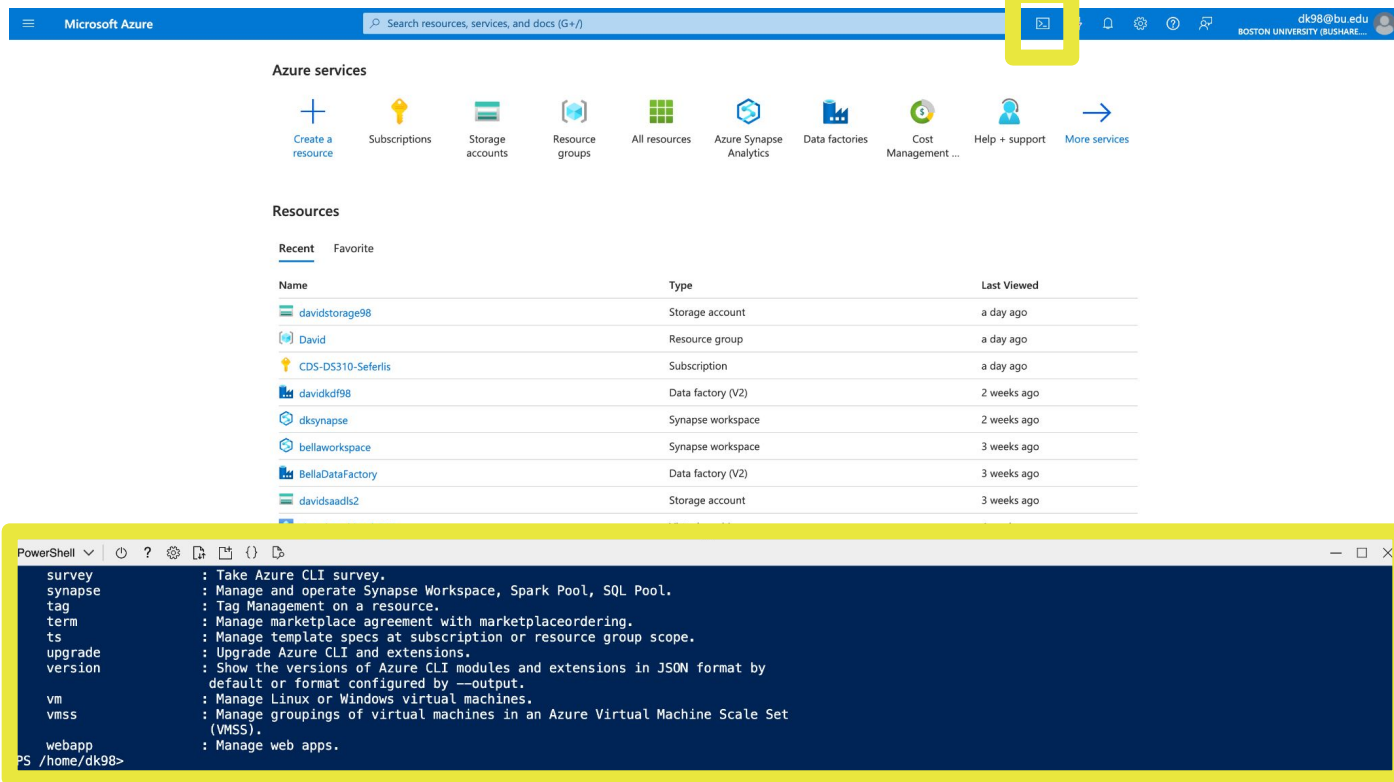


WHY?



\$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. At the top, there's a search bar and a user profile. Below the search bar, there's a section for 'Azure services' with various icons. A yellow box highlights the 'Open Cloud Shell' button, and a yellow arrow points to it. Below this, there's a 'Resources' section with a table of recent resources.

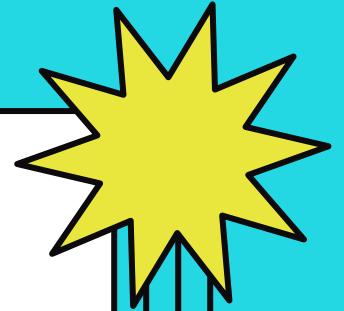
Name	Type	Last Viewed
davidstorage98	Storage account	a day ago
David	Resource group	a day ago
CDS-DS310-Seferlis	Subscription	a day ago
davidkdf98	Data factory (V2)	2 weeks ago
dksynapse	Synapse workspace	2 weeks ago
bellaworkspace	Synapse workspace	3 weeks ago
BellaDataFactory	Data factory (V2)	3 weeks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

Below the resources table, there's a PowerShell terminal window showing 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

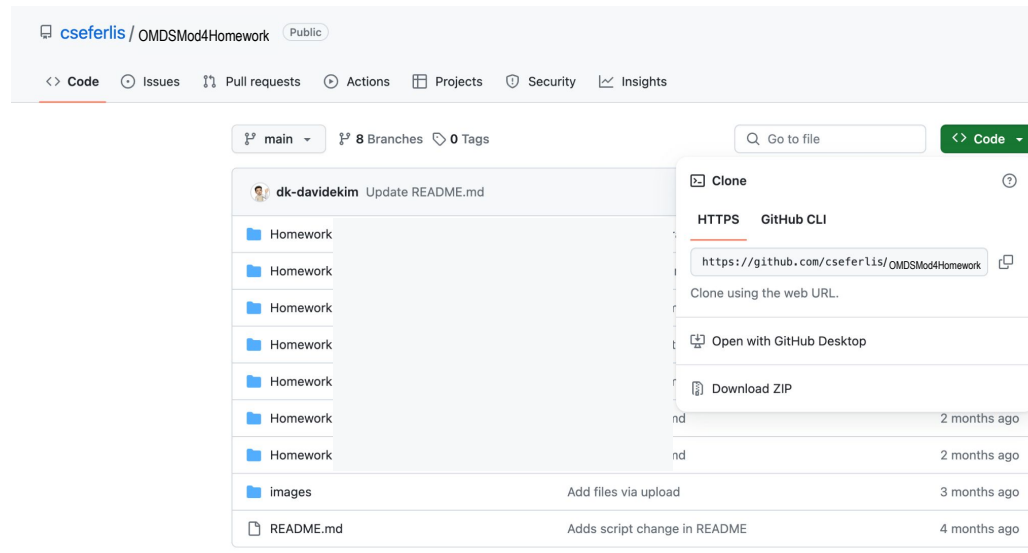
The diagram illustrates the process of navigating into a directory and listing its contents using Linux commands. It consists of three terminal window snippets connected by a large yellow arrow pointing downwards.

Terminal 1: Shows the initial directory as `Downloads`. The command `ls` is entered, resulting in the output `untitled folder`.

Terminal 2: Shows the command `cd 'untitled folder'` being executed, which changes the current directory to `~/Downloads/untitled folder`.

Terminal 3: Shows the command `ls` being executed in the new directory, resulting in the output `helloDS310.txt`.

2. CLONE GITHUB REPO



`https://github.com/cseferlis/OMDSMod4Homework.git`

PS /home/dk98> `git clone https://github.com/cseferlis/OMDSMod4Homework.git`

3. NAVIGATE TO FORMTEMPLATE.SH

OMDSMod4Homework/Homework 1b / Extract And Load /



Adds script for parameters

Name



..



template



README.md



formTemplate.sh

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

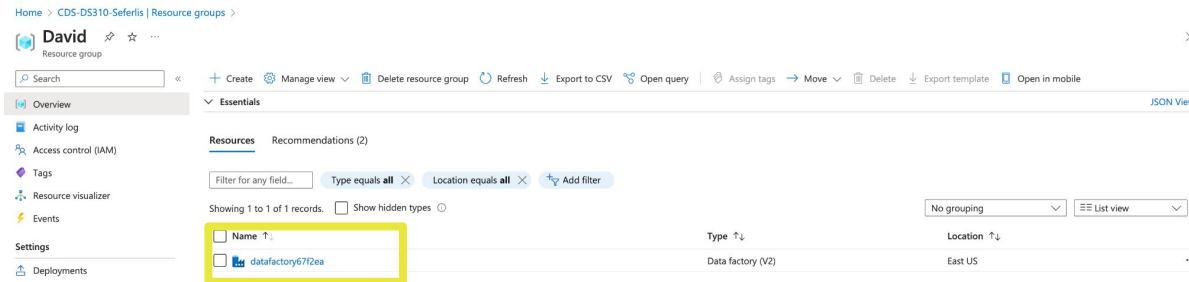
bash ./formTemplate.sh

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

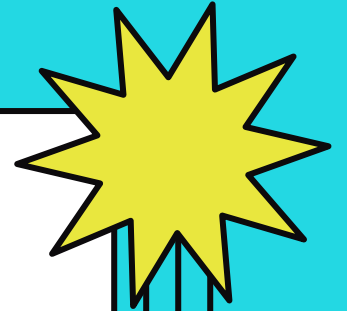


03

Unzip

1. DATA FACTORY

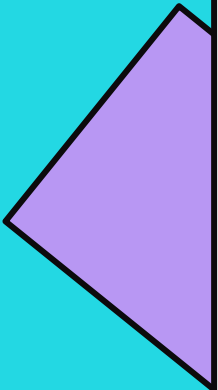
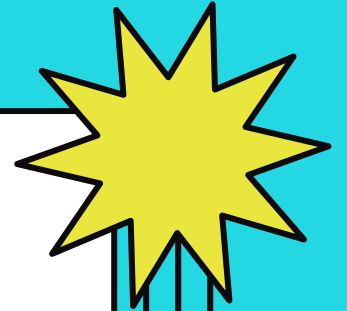
- **Large kitchen** with recipes coming from all over the world
 - You're one of the **Chefs**
 - **Gathering Ingredients** (Data Ingestion)
 - **Preparing Ingredients** (Data Transformation)
 - **Cooking** (Data Computation)
 - **Serving the Dish** (Data Publication)



1. DATA FACTORY

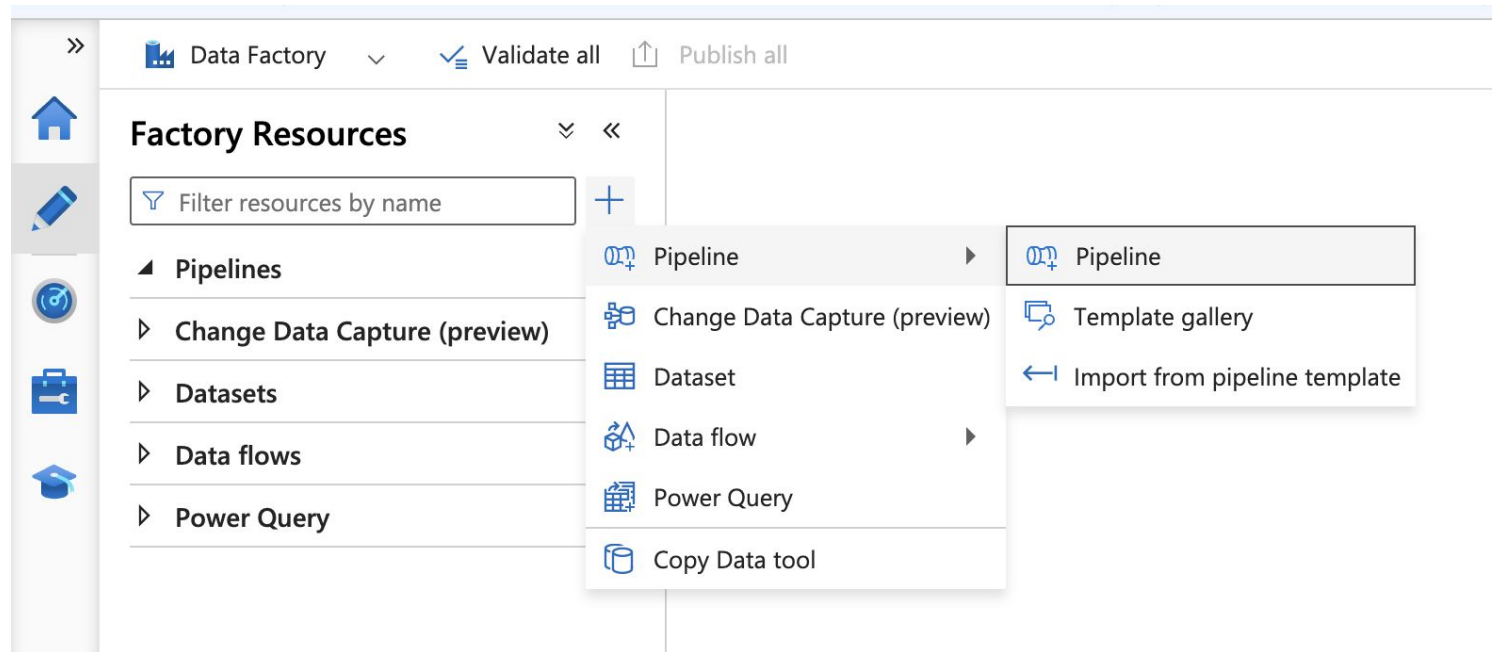
Watch this video at home!

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



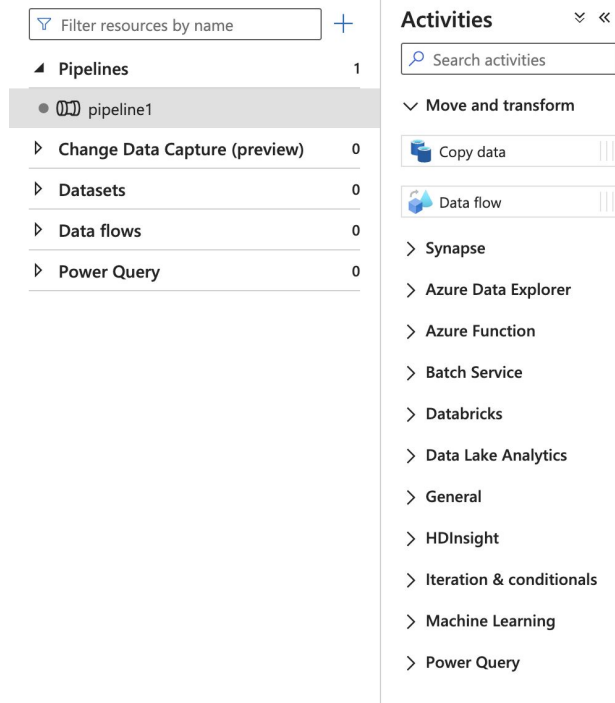
HOMework 1B

Creating a Pipeline



HOMework 1B

Creating Activities



Filter resources by name +

- 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

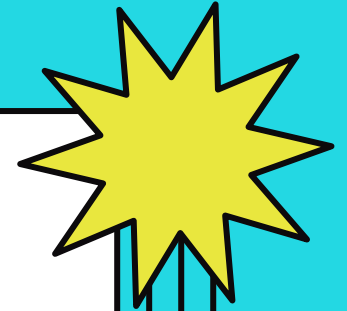
- Set up **Copy Data** activity to **Extract data** from
 - https://static.nhtsa.gov/odi/ffdd/cmpl/FLAT_CMPL.zip
 - Hint: Zip file is a **binary** file!
- 1. Extract it from **HTTP** Linked Service
- 2. Load it to a container in your **Azure Blob Storage**



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 image shows three screenshots from a software interface. The first is a dropdown menu for 'Add trigger' with options 'Trigger now' and 'New/Edit'. The second is a sidebar menu with 'Pipeline runs' highlighted. The third is a page titled 'Activity runs' showing a single run with ID '21c23e6d-83e4-4684-aa20-a0e5562acef8' and status 'Succeeded'.

Add trigger

- Trigger now
- New/Edit

Dashboard

- Home
- Runs
- Pipeline runs**
- 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!

