Moving data into and out of your Docker container and HDFS

    I.   **Moving data into and out of a Docker container**

        1. Moving data directly using the *docker cp* command

The following commands can be used to copy files/folders between a container and the local filesystem. Specifically, the commands below will copy files/folders from the *source_path* to the *destination_path* in the specified *container*. The container does not need to be running.

- `docker cp` *`container:source_path destination_path`*
  - `docker cp optimistic_wright:/me.txt C:\Users\dancikg\Desktop\`
    - (copies the file */me.txt* from the container *optimistic_wright* to the local directory *C:/Users/dancikg/Desktop*)
- `docker cp` *`source_path`* `container:`*`destination_path`*
  - `docker cp C:\file.txt optimistic_wright:/home/`
    - (copies the local file *C:/file.txt* to the */home* directory of the container)

        2. Mount a volume (directory) from your local machine to the container

Mounting a directory makes the directory directly accessible from within the container, as if the directory was part of the container. This is accomplished by using the *-v* option with *docker run*, which allows you to mount a *localDirectory* to a *containerDirectory*, which is a directory within the container. Note that if *containerDirectory* already exists, then its contents will be replaced. In addition, this option only works when creating the container (e.g., it cannot be used with an already running container).

- `docker run -v` *`localDirectory:containerDirectory`*
  - `docker run -it -v C:\Users\dancikg\Desktop:/windows centos bash`
    - (Creates a new container from the *centos* image, and executes the *bash* command. The local directory *C:\Users\dancikg\Desktop* is mounted to the */windows* directory inside of the container.

II. Moving data into and out of HDFS using the command line

To move data from the container to HDFS, commands begin with *hdfs dfs* (formerly *hadoop fs*) and are followed by an option and appropriate arguments. Options and their arguments are similar to their linux counterparts.

Option	Description	Command example	Command explanation
-ls	Lists files and folders in HDFS	`hdfs dfs -ls /user/cloudera/*.txt`	list all .txt files in */user/cloudera/*
-cat	Concatenate and print files in HDFS	`hdfs dfs -cat /file.txt`	Print the contents of *hdfs:/file.txt*
-mv	Moves files and folders within HDFS	`hdfs dfs -mv /file.txt /names.txt`	move *hdfs:/file.txt* to *hdfs:/names.txt*
-cp	Copies files and folders within HDFS (add -f to overwrite)	`hdfs dfs -cp /file.txt /names.txt`	copy *hdfs:/file.txt* to *hdfs:/names.txt*
-put	Puts files and folders into HDFS (add -f to overwrite)	`hdfs dfs -put file.txt /user/cloudera/`	copies *file.txt* from the current directory into *hdfs:/user/cloudera*
-get	Gets files and folders from HDFS and copies to the local filesystem (if the file already exists locally, you must remove it or use a different name)	`hdfs dfs -get /user/cloudera/file.txt .`	copies *file.txt* hdfs:/user/cloudera/file.txt to file.txt in the current local directory
-mkdir	Makes a directory in HDFS (add -p to create intermediate directories as needed)	`hdfs dfs -mkdir /user/cloudera/students`	Makes the directory */user/cloudera/students/* in HDFS

III. Moving data into and out of HDFS using Hue

Hadoop User Experience (Hue) provides a web interface for working with HDFS and Hadoop. When a docker container is running Hue and the container port 8888 is mapped to local port 8888, Hue can be accessed by pointing a web browser to localhost:8888, and logging in using *cloudera* for both the username and password. Clicking the first icon on the top right of the page (hovering over this shows "Manage HDFS") will take you to the HDFS file browser, where you can add, delete, and rename files and directories.