Module 2 Cheat Sheet: Introduction to the Hadoop Ecosystem

Package/Method	Description	Code Example
bin/hadoop	All Hadoop commands are invoked by the bin/hadoop script. Running the Hadoop script without any arguments prints the description for all commands.	Running Hadoop script without arguments: 1. 1 1. bin/hadoop Copied!
cat	Reads each file parameter in sequence and writes it to standard output. If you do not specify a file name, the cat command reads from standard input. You can also specify a file name of - (dash) for standard input.	Create two sample files. 1. 1 2. 2 1. echo "This is file 1" > file1.txt 2. echo "This is file 2" > file2.txt Copied! Use the cat command to read and display the contents of both files 1. 1 1. cat file1.txt file2.txt Copied! Sample output (Contents of file1.txt and file2.txt): 1. 1 2. 2 1. This is file 1 2. This is file 2 Copied!
cd	Used to move efficiently from the existing working directory to different directories on your system.	Basic syntax of cd command: 1. 1 1. cd [options]... [directory] Copied! Example 1: Change directory location to "folder1" 1. 1 1. cd /usr/local/folder1 Copied! Example 2: Get back to the previous working directory 1. 1 1. cd - Copied! Example 3: Move up one level from the present working directory tree 1. 1

Package/Method	Description	Code Example
		1. cd .. Copied!
create table	Used to create a new table in a database	Create a new database (if not already created). 1. 1 1. CREATE DATABASE your_database; Copied! Use the newly created database. 1. 1 1. USE your_database; Copied! Create a new table named "employees" in Hive. 1. 1 2. 2 3. 3 4. 4 5. 5 6. 6 7. 7 8. 8 9. 9 10. 10 1. CREATE TABLE employees (2. id INT, 3. first_name STRING, 4. last_name STRING, 5. email STRING, 6. hire_date DATE 7.) 8. ROW FORMAT DELIMITED 9. FIELDS TERMINATED BY ',' 10. STORED AS TEXTFILE; Copied! Show the list of tables in the database. 1. 1 1. SHOW TABLES; Copied! Sample Output (List of Tables): 1. 1 2. 2 1. OK 2. employees Copied!
curl	A command-line tool (pronounced "curl") that allows data to be exchanged between a device and a server through a terminal. The user specifies the server URL, the location where they want to send	Example 1: Sending a GET request and displaying the response Send a GET request to a server and display the response.

Package/Method	Description	Code Example
	the request, and the data they want to send to the server URL using this command-line interface (CLI).	1. 1 1. curl https://www.example.com Copied! In this example, we use the curl command to send a GET request to https://www.example.com and display the HTML response from the server. ---------- Example 2: Sending data to a server using POST Request: Send a POST request with data to a server and display the response. 1. 1 1. curl -X POST -d "name=John&age=30" https://www.example.com/api Copied! In this example, we use the curl command to send a POST request to https://www.example.com/api with data `name=John&age=30` and display the JSON response from the server.
docker exec	Runs a new command in a running container. It only runs when the container's primary process is running, and it is not restarted if the container is restarted.	Running a command in a running Docker container: Run a new command inside a running Docker container. 1. 1 1. docker exec -it container_name_or_id ls /app Copied! Sample Output (List of files in the '/app' Directory inside the container): 1. 1 2. 2 3. 3 1. file1.txt 2. file2.txt 3. subdirectory Copied! In this example: • `docker exec` is used to run a new command (`ls /app`) inside a running Docker container. • `-it` enables an interactive terminal session, which allows you to see the output of the command. • `container_name_or_id` is the name or ID of the running Docker container you want to execute the command in. • `ls /app` is the command that lists the files and directories in the '/app' directory inside the container.
docker-compose	Compose is a tool for defining and running multi-container Docker applications. It uses the YAML file to configure the services and enables us to create and start all the services from just one configuration file.	Starting Docker containers using docker-compose: Suppose you have a `docker-compose.yml` file like this: 1. 1 2. 2 3. 3 4. 4 5. 5

Package/Method	Description	Code Example
		6. 6 7. 7 8. 8 9. 9 10. 10 1. version: '3' 2. services: 3. web: 4. image: nginx:latest 5. ports: 6. - "80:80" 7. db: 8. image: postgres:latest 9. environment: 10. POSTGRES_PASSWORD: example_password Copied! You can use `docker-compose` to start the services defined in the docker-compose.yml file as follows: Navigate to the directory containing the docker-compose.yml file. 1. 1 1. cd /path/to/your/docker-compose-project Copied! Start the Docker containers defined in the docker-compose.yml file 1. 1 1. docker-compose up Copied!
docker pull	You can download Docker images from the internet.	1. 1 1. docker pull [OPTIONS] IMAGE_NAME[:TAG] Copied!
docker run	It runs a command in a new container, getting the image and starting the container if needed.	1. 1 1. docker run [OPTIONS] IMAGE [COMMAND] [ARG...] Copied!
git clone	You can create a copy of a specific repository or branch within a repository.	1. 1 1. git clone REPOSITORY_URL [DESTINATION_DIRECTORY] Copied!
hdfs dfs	Apache Hadoop hadoop fs or hdfs dfs are file system commands to interact with HDFS. These commands are very similar to Unix commands. Hadoop provides two types of commands to interact with the file system: hadoop fs or hdfs dfs. The major difference is that Hadoop commands are supported with multiple file systems like S3, Azure, and many more.	Example-1: Listing files and directories in HDFS: List files and directories in the root directory of HDFS. 1. 1 1. hdfs dfs -ls / Copied! Example-2: In this example, we use the `hdfs dfs -ls` command to list files and directories in the root directory of HDFS.

Package/Method	Description	Code Example
		1. 1 1. hdfs dfs -ls / Copied! Sample output: 1. 1 2. 2 3. 3 1. drwxr-xr-x - hdfs hduser 0 2023-09-13 10:00 /user 2. drwxrwxrwx - hdfs hduser 0 2023-09-13 10:05 /tmp 3. drwxrwxrwx - mapred hduser 0 2023-09-13 10:10 /mapred Copied! Create a new directory named "mydata" in HDFS. 1. 1 1. hdfs dfs -mkdir /user/your_username/mydata Copied!
hdfs dfs -cat	Display the contents for a file.	Display the contents of a file in HDFS. 1. 1 1. hdfs dfs -cat /path/to/file.txt Copied!
hdfs dfs -mkdir	Creates a directory named path in HDFS	Create a directory in HDFS. 1. 1 1. hdfs dfs -mkdir /user/username/mydirectory Copied!
hdfs dfs -put	Upload a file or folder from the local disk to HDFS.	Upload a file from the local file system to HDFS. 1. 1 1. hdfs dfs -put localfile.txt /user/username/hdfsfile.txt Copied!
LOAD DATA INPATH	Hive provides the functionality to load precreated table entities either from the local file system or from HDFS. This command is used to load data into the hive table.	Load data from HDFS into a Hive table. 1. 1 2. 2 1. LOAD DATA INPATH '/user/username/hdfsfile.txt' INTO TABLE 2. mytable; Copied!
ls	Writes to standard output the contents of each specified Directory parameter or the name of each specified file parameter, along with any other information you ask for with the flags. If you do not specify a file or directory parameter, the ls command displays the contents of the current directory.	Basic command syntax 1. 1 1. ls [options] [file/directory] Copied!

Package/Method	Description	Code Example
		Example 1: Sorts the file names displayed in the order of last modification time. 'r' is for displaying in reverse order 1. 1 2. 2 1. ls -lt 2. ls -ltr Copied! Example 2: Displays hidden files 1. 1 1. ls -a Copied!
mkdir	Used to create one or more directories specified by the Directory parameter. Each new directory contains the standard entries dot (.) and dot dot (..). You can specify the permissions for the new directories with the -m Mode flag.	Create a new directory named "myfolder." 1. 1 1. mkdir myfolder Copied!
SELECT * FROM	Lists all the rows from the table to check if the data has been loaded from the file.	Select all rows from a table. 1. 1 1. SELECT * FROM tablename; Copied!
show tables	Used to see all the tables in the database that have been selected.	Show all tables in the selected database. 1. 1 1. SHOW TABLES; Copied!
tar	Looks for archives on the default device (usually tape) unless you specify another device. When writing to an archive, the tar command uses a temporary file (the /tmp/tar* file) and maintains in memory a table of files with several links.	Create a tar archive of a directory. 1. 1 1. tar -cvf archive.tar /path/to/directory Copied!
wget	Stands for web get. The wget is a free, noninteractive file downloader command. Noninteractive means it can work in the background when the user is not logged in.	Basic syntax of the wget command; commonly used options are `[-V], [-h], [-b], [-e], [-o], [-a], [-q]` 1. 1 1. wget [options]... [URL]... Copied! Example 1: Specifies to download file.txt over HTTP website URL into the working directory. 1. 1 1. wget http://example.com/file.txt Copied!

Package/Method	Description	Code Example
		Example 2: Specifies to download the archive.zip over the HTTP website URL in the background and returns you to the command prompt in the interim. 1. 1 1. wget -b http://www.example.org/files/archive.zip Copied!