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:<br><br>```
bin/hadoop
``` |
| 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.<br><br>```
echo "This is file 1" > file1.txt
echo "This is file 2" > file2.txt
```<br><br>Use the cat command to read and display the contents of bo<br><br>```
cat file1.txt file2.txt
```<br><br>Sample output (Contents of file1.txt and file2.txt):<br><br>```
This is file 1
This is file 2
``` |

Package/Method	Description	Code Example
cd	Used to move efficiently from the existing working directory to different directories on your system.	Basic syntax of cd command: ``` cd [options]... [directory] ``` Example 1: Change directory location to "folder1" ``` cd /usr/local/folder1 ``` Example 2: Get back to the previous working directory ``` cd - ``` Example 3: Move up one level from the present working dire ``` cd .. ```

create table	Used to create a new table in a database	Create a new database (if not already created).
		
		``
		CREATE DATABASE your_database;
		``
		
		Use the newly created database.
		
		``
		USE your_database;
		``
		
		Create a new table named "employees" in Hive.
		
		``
		CREATE TABLE employees (
		    id INT,
		    first_name STRING,
		    last_name STRING,
		    email STRING,
		    hire_date DATE
		)
		ROW FORMAT DELIMITED
		FIELDS TERMINATED BY ','
		STORED AS TEXTFILE;
		``
		
		Show the list of tables in the database.

Package/Method	Description	Code Example
| | | ```
SHOW TABLES;
```<br><br>Sample Output (List of Tables):<br><br>```
OK
employees
``` |
| 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 the request, and the data they want to send to the server URL using this command-line interface (CLI). | Example 1: Sending a GET request and displaying the respon<br><br>Send a GET request to a server and display the response.<br><br>```
curl https://www.example.com
```<br><br>In this example, we use the curl command to send a GET req https://www.example.com and display the HTML response fr server.<br><br>-----------------------<br>Example 2: Sending data to a server using POST Request:<br><br>Send a POST request with data to a server and display the re<br><br>```
curl -X POST -d "name=John&age=30" https://www.example.
```<br><br>In this example, we use the curl command to send a POST re https://www.example.com/api with data `name=John&age=30` a display the JSON response from the server. |
| docker exec | Runs a new command in a running container. It only runs when the container's | Running a command in a running Docker container:<br><br>Run a new command inside a running Docker container.<br><br>```
docker exec -it container_name_or_id ls /app
``` |

Package/Method	Description	Code Example
	primary process is running, and it is not restarted if the container is restarted.	Sample Output (List of files in the '/app' Directory inside the container): ```\nfile1.txt\nfile2.txt\nsubdirectory\n``` In this example: • `docker exec` is used to run a new command (`ls /app`) running Docker container. • `-it` enables an interactive terminal session, which allow see the output of the command. • `container_name_or_id` is the name or ID of the running container you want to execute the command in. • `ls /app` is the command that lists the files and director '/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:
		
		```\nversion: '3'\n services:\n web:\n image: nginx:latest\n ports:\n - "80:80"\n db:\n image: postgres:latest\n environment:\n POSTGRES_PASSWORD: example_password\n```
		
		You can use `docker-compose` to start the services defined in docker-compose.yml file as follows:
		
		Navigate to the directory containing the docker-compose.ym
		
		```\ncd /path/to/your/docker-compose-project\n```
		
		Start the Docker containers defined in the docker-compose.y

Package/Method	Description	Code Example
		`docker-compose up`
docker pull	You can download Docker images from the internet.	`docker pull [OPTIONS] IMAGE_NAME[:TAG]`
docker run	It runs a command in a new container, getting the image and starting the container if needed.	`docker run [OPTIONS] IMAGE [COMMAND] [ARG...]`
git clone	You can create a copy of a specific repository or branch within a repository.	`git clone REPOSITORY_URL [DESTINATION_DIRECTORY]`

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	Example-1:
		
		Listing files and directories in HDFS:
		
		List files and directories in the root directory of HDFS.
		
		`hdfs dfs -ls /`
		
		Example-2: In this example, we use the `hdfs dfs -ls` comm
		list files and directories in the root directory of HDFS.
		
		`hdfs dfs -ls /`
		
		Sample output:
		
		`drwxr-xr-x   - hdfs hduser          0 2023-09-13 10:00`
		`drwxrwxrwx   - hdfs hduser          0 2023-09-13 10:05`

Package/Method	Description	Code Example

	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.	`drwxrwxrwx   - mapred hduser         0 2023-09-13 10:10`
		
		 Create a new directory named "mydata" in HDFS.
		
		 ```hdfs dfs -mkdir /user/your_username/mydata```

hdfs dfs -cat	Display the contents for a file.	Display the contents of a file in HDFS. ```hdfs dfs -cat /path/to/file.txt```
hdfs dfs -mkdir	Creates a directory named path in HDFS	Create a directory in HDFS. ```hdfs dfs -mkdir /user/username/mydirectory```
hdfs dfs -put	Upload a file or folder from the local disk to HDFS.	Upload a file from the local file system to HDFS. ```hdfs dfs -put localfile.txt /user/username/hdfsfile.txt```
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	Load data from HDFS into a Hive table. ```LOAD DATA INPATH '/user/username/hdfsfile.txt' INTO TABLE mytable;```

Package/Method	Description	Code Example
	data into the hive table.	

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
		
		```ls [options] [file/directory]```
		
		Example 1: Sorts the file names displayed in the order of last modification time. 'r' is for displaying in reverse order
		
		```ls -lt```
		```ls -ltr```
		
		Example 2: Displays hidden files
		
		```ls -a```

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	Create a new directory named "myfolder." ```mkdir myfolder```

Package/Method	Description	Code Example
	for the new directories with the -m Mode flag.	
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. ``` SELECT * FROM tablename; ```
show tables	Used to see all the tables in the database that have been selected.	Show all tables in the selected database. ``` SHOW TABLES; ```
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. ``` tar -cvf archive.tar /path/to/directory ```
wget	Stands for web get. The wget is a free, noninteractive	Basic syntax of the wget command; commonly used options V], [-h], [-b], [-e], [-o], [-a], [-q] ``` wget [options]... [URL]... ```

Package/Method	Description	Code Example

	file downloader command. Noninteractive means it can work in the background when the user is not logged in.	Example 1: Specifies to download file.txt over HTTP website the working directory.
		
		`wget http://example.com/file.txt`
		
		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.
		
		`wget -b http://www.example.org/files/archive.zip`

Changelog

Date	Version	Changed by	Change Description
2023-09-20	2.0	Kunal Merchant	QC Reviewed
2023-09-20	1.0	Gagandeep Singh	Initial version created