HDFS File System Operations

These commands are used to interact with the **Hadoop Distributed File System (HDFS)**, which is the storage layer for Hadoop.

- **`hdfs dfs -ls /path/`** Lists the files and directories in the specified HDFS path. This is used extensively to verify the existence of input files and check the contents of output directories after running jobs.
- **`hdfs dfs -mkdir /path/`** Creates a new directory within HDFS.
- **`hdfs dfs -chown user:group /path/`** Changes the **owner and group** of a file or directory in HDFS.
- **`hdfs dfs -put local_file hdfs_path`** Copies a file from the local Linux file system (e.g., `/usr/mybigdata/wordscount.txt`) into HDFS, making it available for Big Data processing.
- **`hdfs dfs -cat /path/file` Prints the content** of a file located in HDFS to the terminal, often used to inspect the result files (`part-r-00000`) produced by MapReduce jobs.
- **`hdfs dfs -rm -r /path/` Deletes a file or directory** in HDFS. The `-r` flag indicates a recursive deletion, which is necessary when deleting directories (like old job outputs).

MapReduce Job Execution

These commands initiate and run Big Data processing jobs using the Hadoop framework.

- **`hadoop jar JAR_file MainClass input_path output_path`** Executes a **MapReduce job** packaged in a Java Archive (`.jar`) file (e.g., `WordCount.jar`). This command specifies the program to run, the input data location in HDFS, and the HDFS location where the results should be written.
- **`hadoop jar Exercise6-1.0.0.jar ...`** Similar to the above, this executes a specific, custom-developed MapReduce application (`Exercise6...jar`), indicating the user is working on different data analysis tasks (like processing `sensors.csv`).

Local File System and Session Control

These are standard Linux shell commands used for navigation and user environment management.

- **`cd /path/` Changes the current working directory** in the local Linux file system.
- **`ls`** Lists the files and directories in the **current local directory**.
- **`whoami`** Shows the effective **username** of the current user.
- **`sudo su username`** Switches the current user to the specified `username` (e.g., `hadoop`), typically requiring administrative privileges.
- **`cd ..`** Moves the current working directory **up one level** in the local file system hierarchy.