diff is a command-line utility that **allows you to compare two files line by line**. It can also compare the contents of directories. The diff command is most commonly used to create a patch containing the differences between one or more files that can be applied using the patch command.

diff stands for **difference**. This command is used to display the differences in the files by comparing the files line by line. Unlike its fellow members, [cmp](https://www.geeksforgeeks.org/cmp-command-linux-examples/) and [comm](https://www.geeksforgeeks.org/comm-command-linux-examples/), it tells us which lines in one file have is to be changed to make the two files identical.

The important thing to remember is that **diff** uses certain **special symbols** and **instructions** that are required to make two files identical. It tells you the instructions on how to change the first file to make it match the second file.

**Special symbols are:**

**a : add**

**c :** **change**

**d :** **delete**

**Syntax :**

**diff [options] File1 File2**

Lets say we have two files with names **a.txt** and **b.txt** containing 5 Indian states.

**$ ls**

a.txt b.txt

**$ cat a.txt**

Gujarat

Uttar Pradesh

Kolkata

Bihar

Jammu and Kashmir

**$ cat b.txt**

Tamil Nadu

Gujarat

Andhra Pradesh

Bihar

Uttar pradesh

Now, applying **diff** command without any option we get the following output:

**$ diff a.txt b.txt**

0a1

> Tamil Nadu

2,3c3

< Uttar Pradesh

Andhra Pradesh

5c5

Uttar pradesh

Let’s take a look at what this output means. The first line of the **diff** output will contain:

* Line numbers corresponding to the first file,
* A special symbol and
* Line numbers corresponding to the second file.

Like in our case, **0a1** which means **after** lines 0(at the very beginning of file) you have to add **Tamil Nadu** to match the second file line number 1. It then tells us what those lines are in each file preceded by the symbol:

* Lines preceded by a **<** are lines from the first file.
* Lines preceded by **>** are lines from the second file.
* Next line contains **2,3c3** which means from line 2 to line 3 in the first file needs to be changed to match line number 3 in the second file. It then tells us those lines with the above symbols.
* The three dashes **(“—“)** merely separate the lines of file 1 and file 2.

As a summary to make both the files identical, first add *Tamil Nadu* in the first file at very beginning to match line 1 of second file after that change line 2 and 3 of first file i.e. *Uttar Pradesh* and *Kolkata* with line 3 of second file i.e. *Andhra Pradesh*. After that change line 5 of first file i.e. *Jammu and Kashmir* with line 5 of second file i.e. *Uttar Pradesh*.

Now let’s see what it looks like when **diff** tells us that we need to delete a line.

**$ cat a.txt**

Gujarat

Andhra Pradesh

Telangana

Bihar

Uttar pradesh

**$ cat b.txt**

Gujarat

Andhra Pradesh

Bihar

Uttar pradesh

**$ diff a.txt b.txt**

3d2

< Telangana

Here above output **3d2** means delete line 3rd of first file i.e. *Telangana* so that both the files **sync up** at line 2.

**Options**

Linux system offers two different ways to view the **diff** command output i.e. **context mode** and **unified mode**.

**1. -c (context) :** To view differences in context mode, use the **-c** option. Lets try to understand this with example, we have two files **file1.txt** and **file2.txt**:

**$ cat file1.txt**

cat

mv

comm

cp

**$ cat file2.txt**

cat

cp

diff

comm

**$ diff -c file1.txt file2.txt**

\*\*\* file1.txt Thu Jan 11 08:52:37 2018

--- file2.txt Thu Jan 11 08:53:01 2018

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\* 1,4 \*\*\*\*

cat

- mv

- comm

cp

--- 1,4 ----

cat

cp

+ diff

+ comm

The first file is indicated by **\*\*\***, and the second file is indicated by **—**.   
The line with **\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*** is just a separator.

The first two lines of this output show us information about **file 1** and **file 2**. It lists the file name, modification date, and modification time of each of our files, one per line.   
The next line has three asterisks **\*\*\*** followed by a line range from the first file (in our case lines 1 through 4, separated by a comma). Then four asterisks **\*\*\*\***. After that it shows the contents of the first file with the following indicators:

**(i)** If the line needs to be unchanged, it is prefixed by two spaces.   
**(ii)** If the line needs to be changed, it is prefixed by an symbol and a space. The symbol means are as follows:

**(a) + :** It indicates a line in the second file that needs to be added to the first file to make them identical.   
**(b) – :** It indicates a line in the first file that needs to be deleted to make them identical.   
Like in our case, it is needed to delete *mv* and *comm* from first file and add *diff* and *comm* to the first file to make both of them identical.

After that there are three dashes **—** followed by a line range from the second file (in our case lines 1 through 4, separated by a comma). Then four dashes **—-**. Then it shows the contents of the second file.

**2. -u (unified) :** To view differences in unified mode, use the **-u** option. It is similar to context mode but it **doesn’t display any redundant information** or it shows the information in concise form.

**$ cat file1.txt**

cat

mv

comm

cp

**$ cat file2.txt**

cat

cp

diff

comm

**$ diff -u file1.txt file2.txt**

--- file1.txt 2018-01-11 10:39:38.237464052 +0000

+++ file2.txt 2018-01-11 10:40:00.323423021 +0000

@@ -1,4 +1,4 @@

cat

-mv

-comm

cp

+diff

+comm

The first file is indicated by **—**, and the second file is indicated by **+++**.   
The first two lines of this output show us information about **file 1** and **file 2**. It lists the file name, modification date, and modification time of each of our files, one per line.   
After that the next line has two at sign **@** followed by a line range from the first file (in our case lines 1 through 4, separated by a comma) prefixed by **–** and then space and then again followed by a line range from the second file prefixed by **+** and at the end two at sign **@**. Followed by the file content in output tells us which line remain unchanged and which lines needs to added or deleted(indicated by symbols) in the **file 1** to make it identical to **file 2**.

**3. -i :** By default this command is *case sensitive*. To make this command *case in-sensitive* use **-i** option with **diff**.

**$ cat file1.txt**

dog

mv

CP

comm

**$ cat file2.txt**

DOG

cp

diff

comm

Without using this option:

**$ diff file1.txt file2.txt**

1,3c1,3

< dog

< mv

DOG

> cp

> diff

Using this option:

**$ diff -i file1.txt file2.txt**

2d1

diff

**4. –version :** This option is used to display the version of **diff** which is currently running on your system.

$ diff --version

diff (GNU diffutils) 3.5

Packaged by Cygwin (3.5-2)

Copyright (C) 2016 Free Software Foundation, Inc.

License GPLv3+: GNU GPL version 3 or later .

This is free software: you are free to change and redistribute it.

There is NO WARRANTY, to the extent permitted by law.

Written by Paul Eggert, Mike Haertel, David Hayes,

Richard Stallman, and Len Tower.