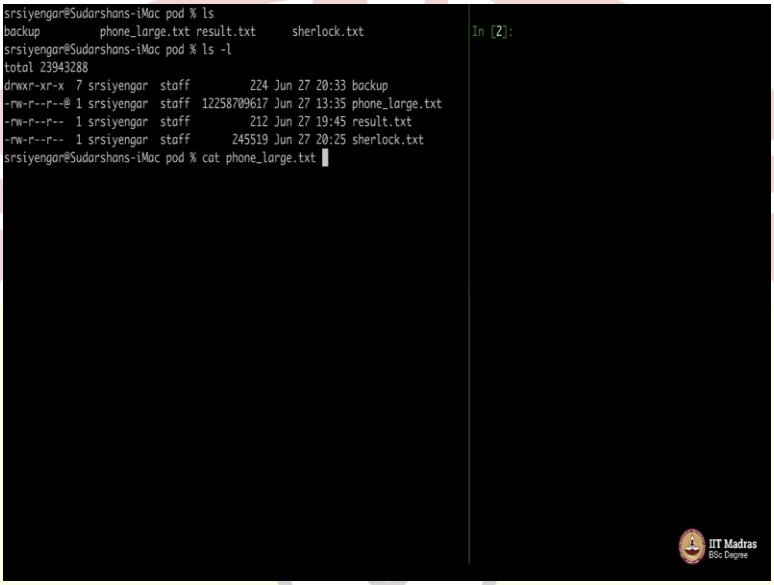
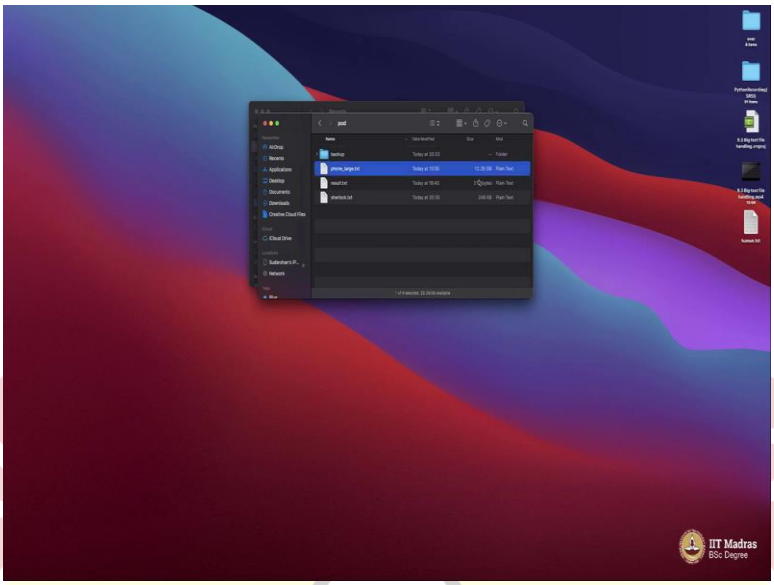

# IIT Madras

## ONLINE DEGREE

**Programming in Python**
**Professor. Sudarshan Iyengar**
**Department of Computer Science & Engineering**
**Indian Institute of Technology, Ropar**
**Very big files a tip**

(Refer Slide Time: 0:16)

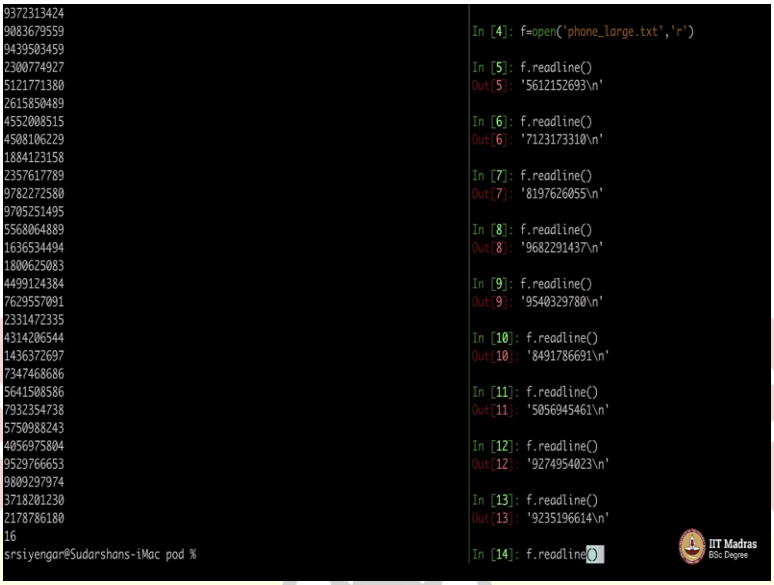


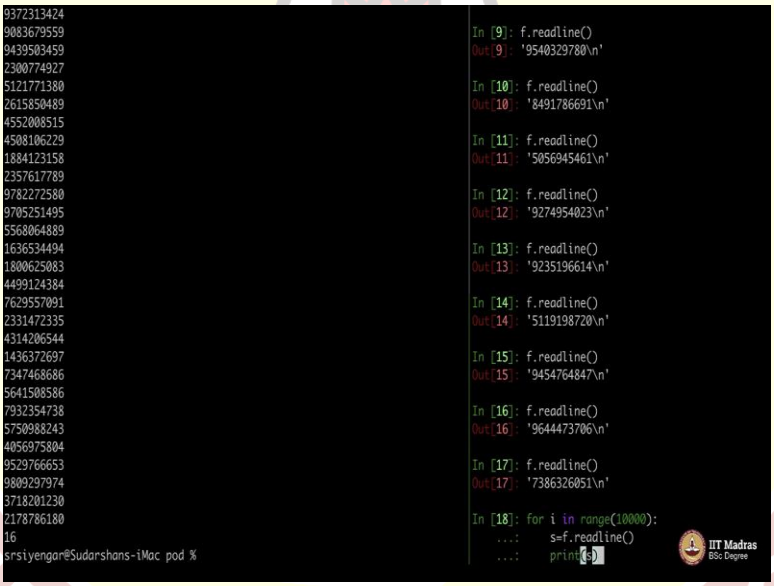




So, I have this big file here, let us see, it is called phone large. It contains a huge list of phone numbers. So, let us see how much. This, up to this point is 8 million, 58 million, 258 million, 2 billion, 12 billion. This is 12 GB of data. And that is a whole lot. I cannot even imagine. If I display this, it will take, I mean, hours together for it to get even over. And if you want to search for a phone number here in this huge database, it is going to be very difficult. In fact, let me try opening this on my desktop. It is not going to be easy. I will show you.

So, you see, I will try to open this file, trying to open this file. It is here phone large 12 GB, you see. As I double click this, you see it is stuck. It does not move. It cannot open basically. In no editor can you open this. Let me close this and then get back. So, it close up and it never gets over. I just cut it short by control c. Meaning, I mean, I broke it, broke the process. And now, how do I go through this particular file for an element?

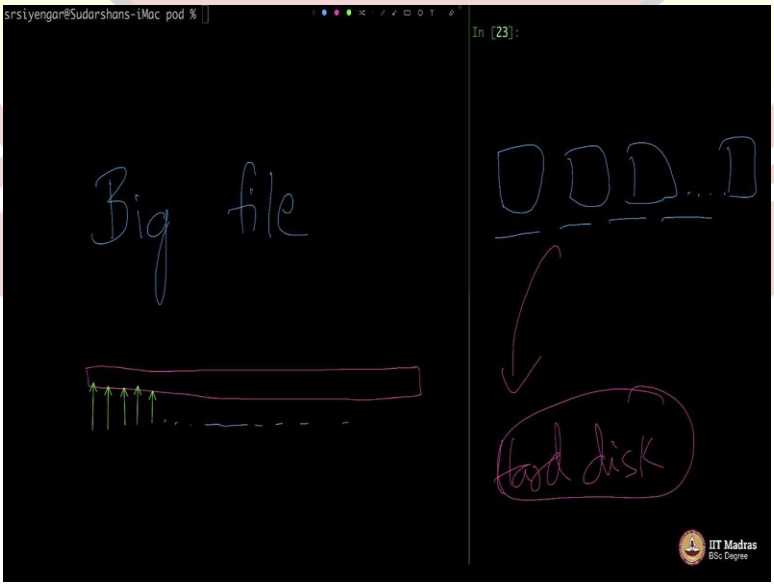
(Refer Slide Time: 1:56)




So, what I can do is, I can indeed, let me see where I am. So, what I can do is, f equals open, what was that, phone large dot txt, that is the name of the, I read from it. So, the best part is, no matter how big the file is. I repeat, no matter how big the file is, you can always read from it. The moment you say read line, it takes you to the next line, next line, next line, next line, and so on. For i in range, let us say, 10,000 lines, if you want to read, read line. Let me append this to some s and print s. See, this is always possible, and you can keep doing it, the next 10,000, the next 10,000, the next 10,000.

(Refer Slide Time: 2:47)

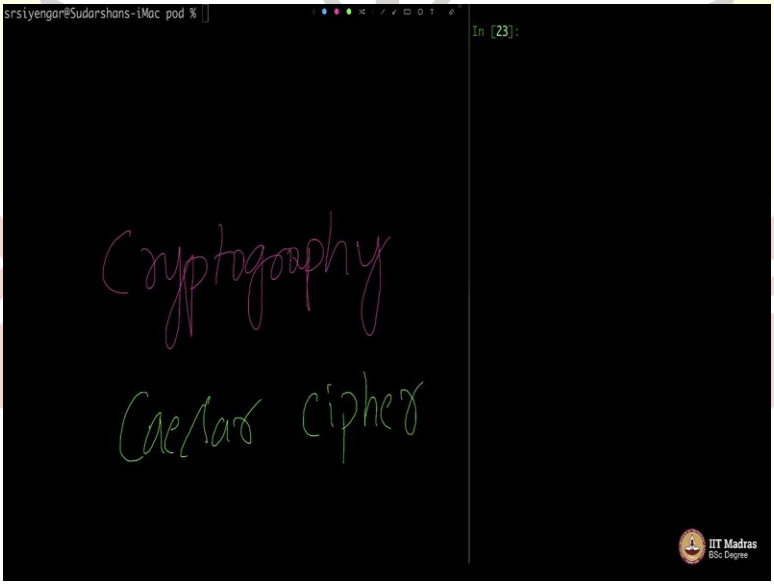


So, what do I mean by this little illustration? By this, I mean, irrespective of how big the file is, take a big file, it may not open, but you can go line by line. You can basically go line by line by line by line of the file. I hope it is clear to you. It might take a long time for it to finish, but it will never get hung. Why is that? That is because your file system. So, let me illustrate that. Your file system is like a big tape like this. It is a big tape. It will start seeing what is in the beginning of this second, third, fourth, fifth and so on like this, and so on. It can always go linearly in this fashion, in your file one after the other.

So, although the file is 12 GB, it will take its own time, but it will execute. I mean, it will go through it one by one. That is the advantage of a big file. That is how your movie files work. It is, movies are generally it is a list of pictures. There will be millions and millions of billions of pictures. And when it displays the pictures in a sequence, it appears like a movie.

And that is taken from your hard disk and processed one file at a time. That is also another program, you see. Your movie players are simply yet another program written in some programming language. So, it is important for you to understand that file handling is possible irrespective of how big the file is just that it will take a long time to process it.

(Refer Slide Time: 4:54)




So, now let me go ahead and show you a nice illustration of this concept from cryptography, which we discussed already called the Caesar cipher that will be our next program.