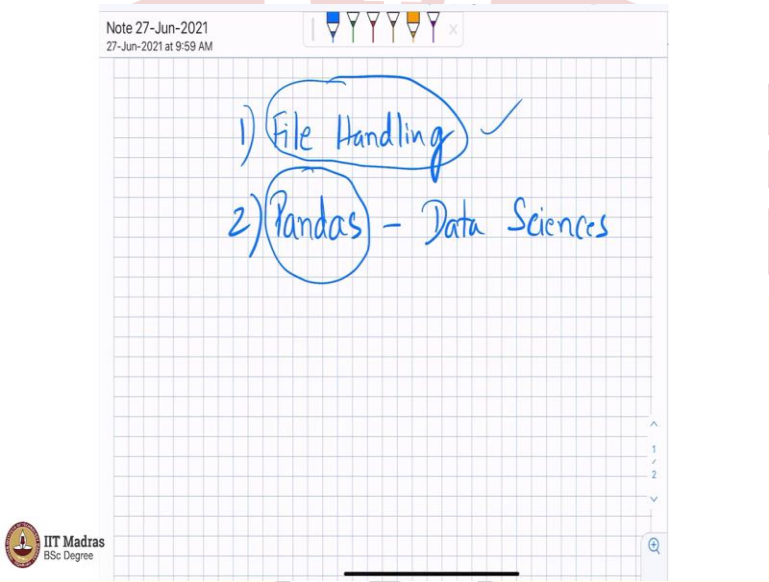
IIT Madras

ONLINE DEGREE

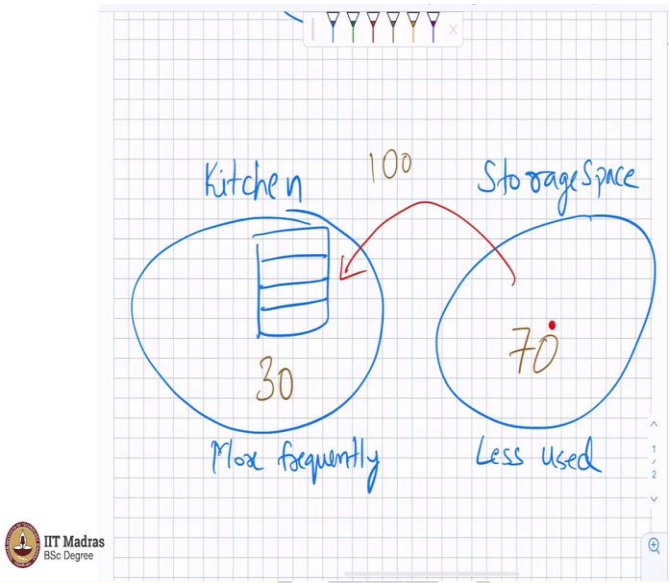
Programming in Python
Professor Sudarshan Iyengar
Department of Computer Science & Engineering
Indian Institute of Technology, Ropar
Introduction to File Handling

(Refer Slide Time: 00:16)



In this week's lecture, we will be looking at two things, one being file handling, and what is called the pandas library function. This is very important for people who would like to take up data sciences. This is a, sort of a clever way and an easy way to handle big files. But the basics of file handling will be taught in the first part of this week's lecture. So, what is file handling and why does one care about it?

(Refer Slide Time: 00:54)

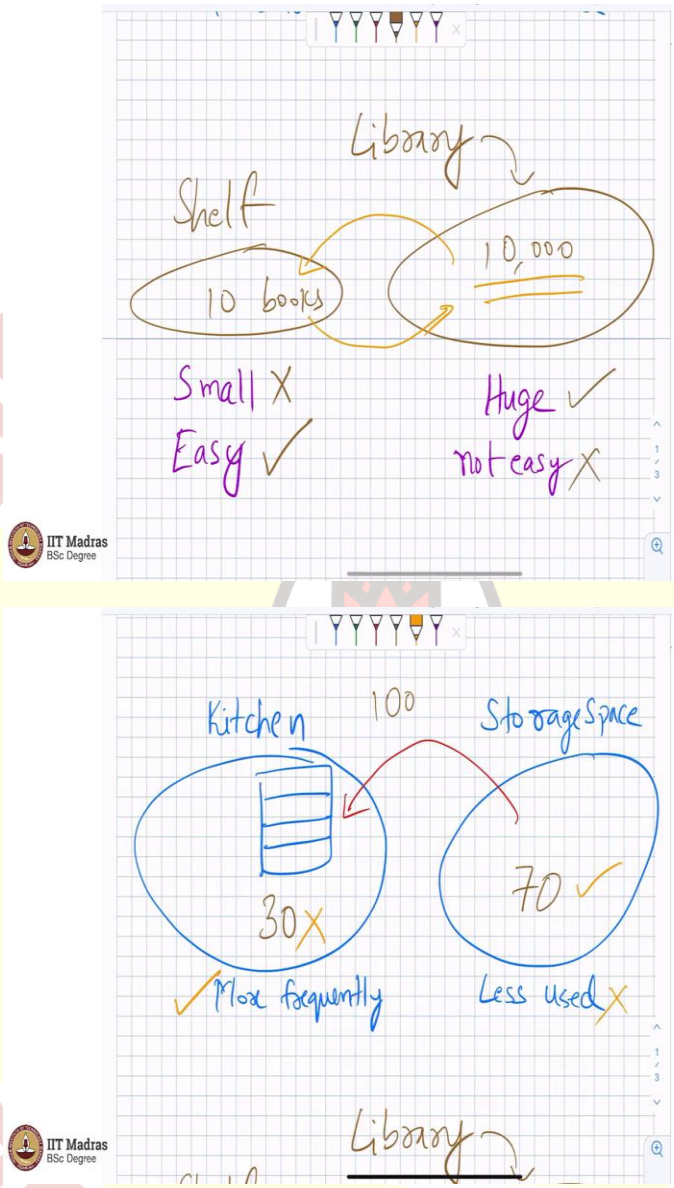


So, let us start with a couple of examples. Example number one, assuming, so let us take for example, your kitchen in your house. It is a small space with maybe a shelf where you keep all your vessels. And there is also some storage space outside your kitchen where you keep stuff that is less frequently used.

And in the kitchen, you keep more frequently used things. So, why, my question is, why will not you keep this entire thing here itself in the kitchen. For example, you have some 100 vessels, you have 100 vessels, out of which you keep only 30 vessels here, and there are 70 less used vessels in some storage space.

Why would you do that? Why cannot we keep all 100 here, that is because in kitchen, we may not have that much of space. And if you make space for 100 vessels, 70 of which is seldom used, you will be using bigger chunk of a kitchen just to store something that is never used, you see. So, let us used stuff goes here, more frequently used stuff comes here and also you keep using it very often. So, taking it from the shelf, putting it back to the shelf is ordeal if you keep doing it from the storage space. I think it is very clear to you. This is analogously evident in many works of our life.

(Refer Slide Time: 02:35)

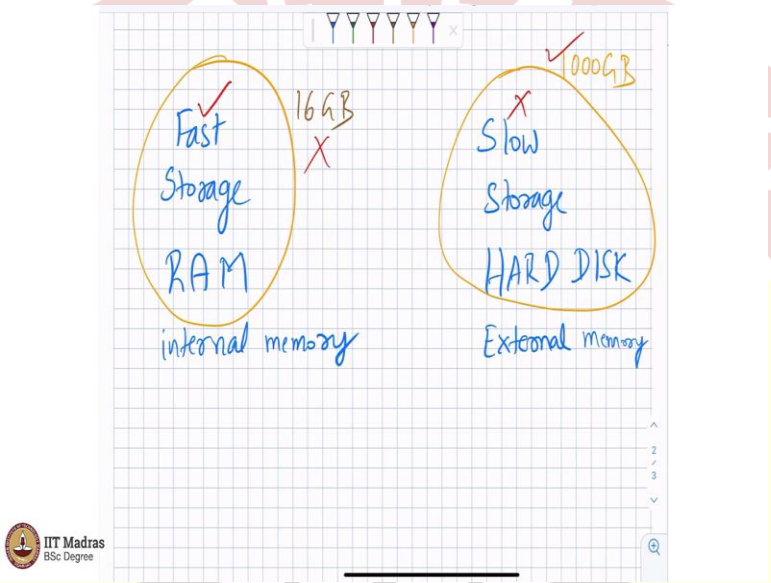


For instance, if I were to consider another example quickly, the example of library, you get some, you have your own shelf, where you keep some 10 books and you have a library with around 10,000 books. Whenever you want, you get books from the library and keep it here and then return it back after using it. Why cannot you house all 10,000 books in your own house? Of course, you cannot do that. I mean, you do not have so much space. But then library comes. So, you see, let us write down the advantage and disadvantage of this.

The advantage of the library is it is huge, not very easy to access, not easy to access. This, your shelf is small, but easy to access, you see. As and always life is full of trade-offs. You have a

nice thing here, but not so nice thing here. It is small. You have a nice thing here. It is a huge repository. It is not very easy to access. Exactly they are similar to what we saw here, where it is used more frequently advantage, but it is very small disadvantage. It is really a big advantage, and it is difficult to use frequently.

(Refer Slide Time: 04:02)



So, very similarly, even on your computer, you have two different types of storage. One is a fast storage, other one is a slow storage. What do you mean by slow storage? The fast storage is basically what is called the RAM also called internal memory. This is basically your hard disk which is actually a tad bit slow, in fact, a lot slower than your RAM. And this is your external memory.

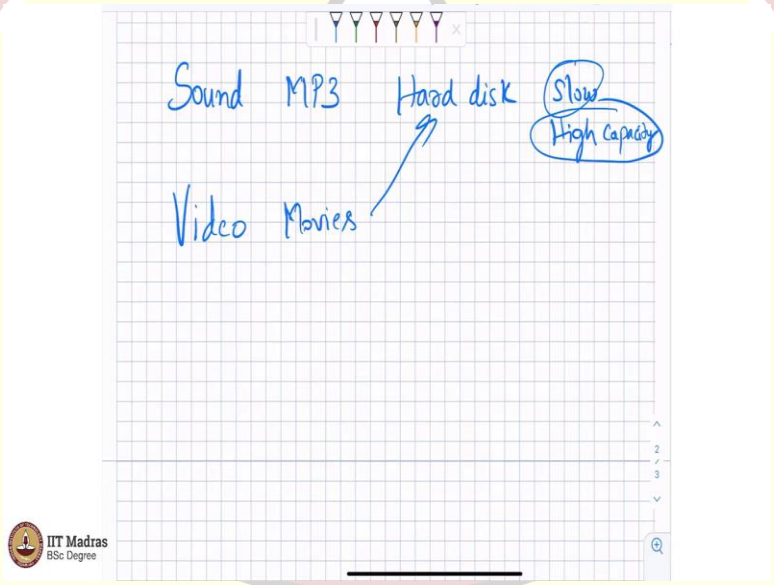
There is a reason why they call it internal and external. Internal because it is very much on your sort of motherboard, while external memory can be, it is a hard disk. It can be even outside your computer connected from the USB port or in a way, it can also be in your CPU. Most of the times you have your internal hard disk, that is also called internal. I mean, do not get confused with the word internal, external. It is basically from 60-70s.

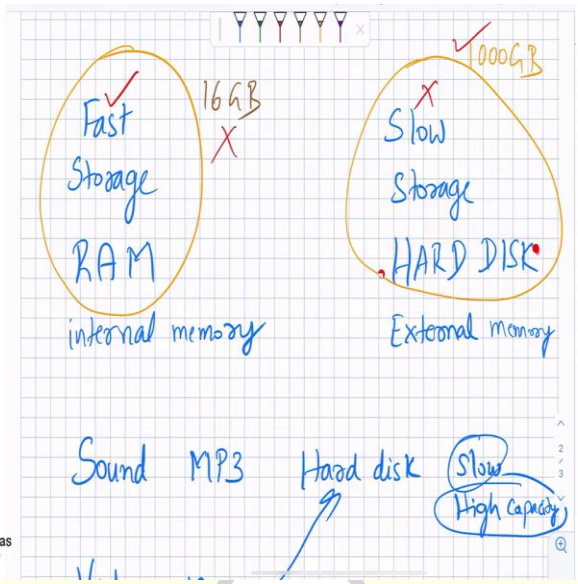
So, internal always meant random access memory, external meant hard disk, internal to the motherboard and external to the motherboard. Anyway, you also connect hard disk to the motherboard in some way. But so you get the point. So, let us not break your ahead much about it. The other, the basic idea here is that RAM is really fast. But you get to store something like

most of us have 16 GB of RAM. While as, while if you look at hard disk, we have some 1 terabyte, which is 1000 GB, 1024 GB to be precise, roughly 1000 GB.

So, as you can see, this is small. This is small disadvantage. This is big advantage. But this is fast advantage. This is slow disadvantage. You cannot have a lot and keep it fast. That is not possible. You cannot have it fast and keep it high. That is not possible. So, this is again, the trade offs in life that we see as I keep saying bigger the vehicle, lesser the fuel efficiency, smaller the vehicle more the fuel efficiency, but the vehicle is small. It is only a two seater, while a bus is a 60 seater.

(Refer Slide Time: 06:47)



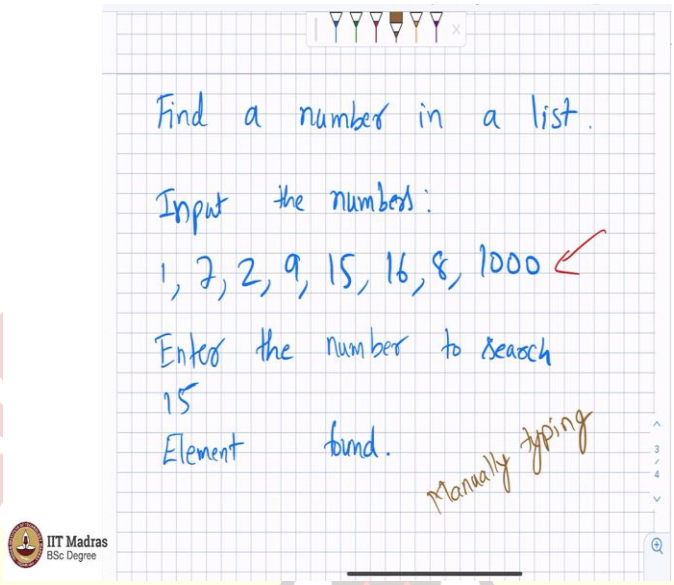


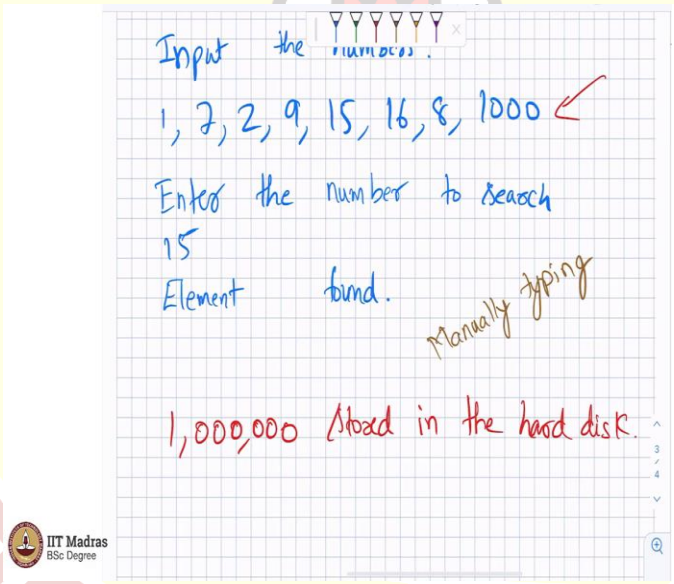
So, with this, why do we even care about this? With this, I will go ahead with the motivation for this external storage device, like a hard disk. So, why would, why do we even care for it? Let us say a small sound file, also called the mp3 file, the music that you keep listening from your computer or from your mobile phone, it is stored in your external hard disk, which is actually slow, but high in capacity. Even your video files, movies, etc. also called the, it is generally in the mp4 format or the raw avi format, whatever technical jargons aside, that is also stored in hard disk and that is also slow, but in high capacity.

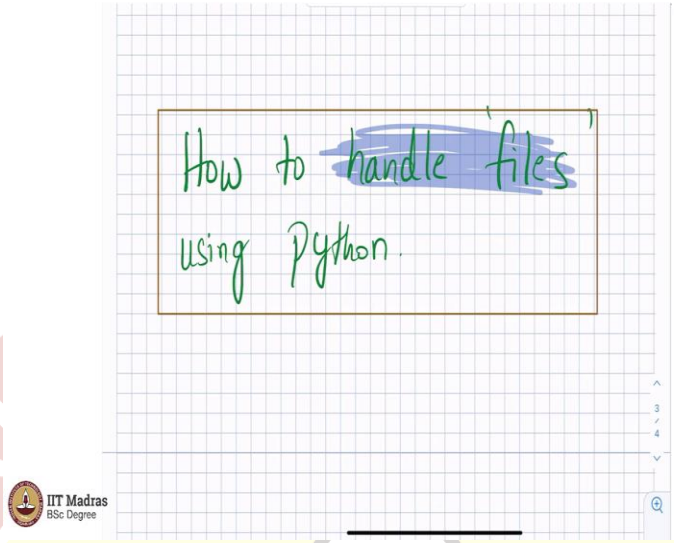
You can always wonder, why do not we store movies in RAM, because it is very fast. So, you cannot, simply because the storage is very less. What if you say, I have a small movie which is only 1 GB, which I can store here. There is another disadvantage of this RAM. The moment you switch off the computer, whatever is in RAM goes away. While whatever is in your hard disk stays forever, almost forever. Unless there is some problem with the disk it is always there in the hard disk for years to come.

So, some of these sounds and video files, there is no other go, we have to use a hard disk only. I think most of these things are obvious. As I am stating it, you are probably wondering, these are some kindergarten syllabus of computer science. We all know it. Good, if you know it.

(Refer Slide Time: 08:30)







Let us go ahead and motivate the reason why we are discussing this in a course like Python. Let me take a typical example of find a number in a list. So, what do you do? Input, you will say input the numbers and then I will put the numbers. I will say 1, 7, 2, 9, 15, 16, 8, 1000. And then you will say enter the number to search. But I will say 15 and you will display your program basically this is the terminal that is displaying the program. I am doing it the lousy way on this let. Enter the number to search, I say 15, and you will say element not found.

Just pause for a second and see what is happening here. Print, input the numbers, input 1, 7, 2, 9, 15, 16, 18, 8, 1000. You take this onto a list and you search through the list if 15 is there or not. If it is there, you say found. I am sorry, it should be found, element found, not element not found. So, you say found. If it is not there, you say element not found, as simple as that. But then pause for a minute and think, where is this coming from? Where is this element coming from? You are manually typing it.

What if you have, let us say, 1 million numbers, you cannot manually type it, stored in the hard disk. How will you search for the presence of an element there amongst these 1 million numbers? There has to be a way to do it. You know how to do it over a list. But if it is on a hard disk, how will you search for and that is precisely what we will be discussing in how to handle files using Python.

We will discuss this in detail now. I will open the terminal now and I will tell you, how do you take a file from the hard disk, work over it and then present your output or do your analysis or

whatever. How to handle files in the first place is what we will be discussing in this week's first section. So, let us go ahead, switch on a terminal and try to see how we can handle files.