

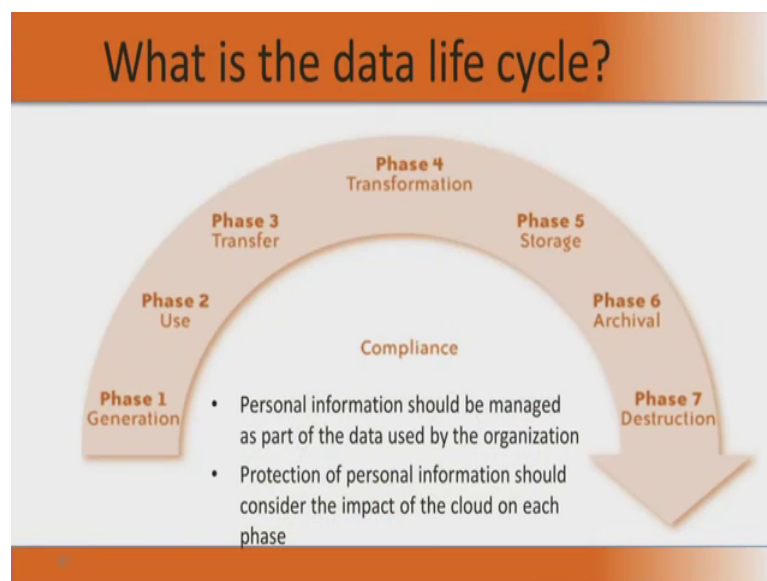
Introduction to Information Security
Prof. V. Kamakoti
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture - 27

So, the ultimate objective is to go and protect data, right because data is when process becomes information. I take you, all to the first module. So, raw data when processed in becomes information. So, the ultimate objective is that data needs to be protected and it should not be corrupted. And so, then when we want to study cloud. So, who manipulates data, it is not the operating system it is the application.

So, the application manipulates data. So, when we look at application level security we should look at what do you mean by application level security? I need to go and protect the data that is manipulated by the application. So, to understand that in a good perspective, we need to look at the data life cycle. What happens to the data? The data is born and it dies.

(Refer Slide Time: 01:17)



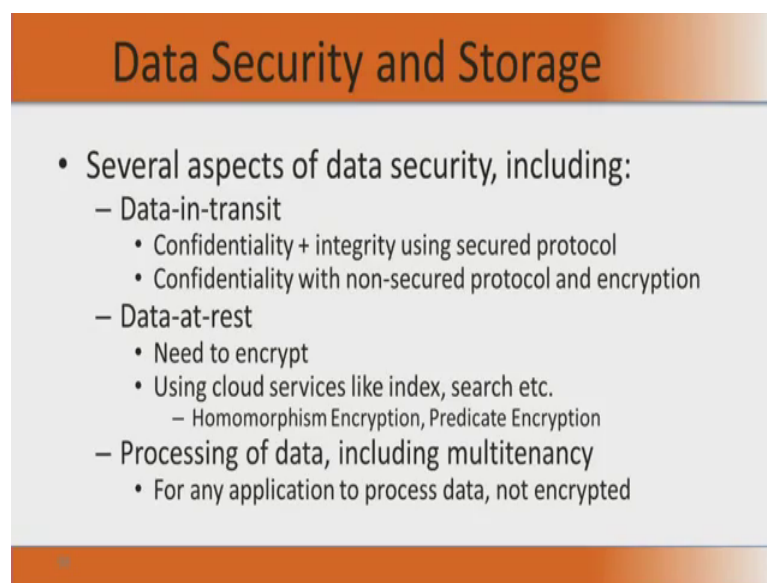
So, between the birth and it is death it should be kept clear. The necessary confidentiality for the data should be there, integrity of the data should be there and availability of the data whenever it is required by some application should also be there. Now, when we move from a data center to a cloud, what would be the impact of this cloud on the different phases of this data, is what is to be studied.

Now, as you seen the slide you have the data first it is generated, then it is used then it

can be transferred, then it can be transformed, then it can be stored, then it can be archival and finally, it can be destructed. So, at least there are seven phases in the data life in the data life cycle. What is the impact of moving from a pure data center model into a cloud model and what would be the impact of this movement on each of these phases if you go and study

in detail from a security point of view that makes that will give us much more of what we are trying to secure, when we do the computations through our applications. And that is what we will see in this session.

(Refer Slide Time: 02:52)



The slide features a title 'Data Security and Storage' in white text on an orange gradient background. Below the title, on a light gray background, is a bulleted list of data security aspects. The list includes 'Data-in-transit' with sub-points for confidentiality and integrity using secured protocols and confidentiality with non-secured protocols and encryption. It also includes 'Data-at-rest' with sub-points for the need to encrypt and the use of cloud services like index and search, with a further sub-point for Homomorphism Encryption and Predicate Encryption. The final point is 'Processing of data, including multitenancy', with a sub-point for any application to process data, not encrypted.

Now, when we look at data security and storage, of course there is something like the data when it is generated it can be in transit. That it is moving from one location physical location to another physical location. And when that movement happens we need to ensure that confidentiality and integrity is maintained. And we do that using a secure protocol. So, we may also do it with the non secure protocol, but with an encryption.

So, any one of this is possible and when the data is at rest, that is it is located in a storage. In a data center within a data center though not advisable we can store it in some un encrypted form, but when we have putting it to on a cloud type of infrastructure it is absolutely necessary. Though the cloud service as infrastructure can say that they will provide necessary isolation, it is in the best interest to keep it in an encrypted form. Now, when we store data in an encrypted form, then some of the known services right.

We do not we do not write our software, but there can be services on the cloud like

indexing, searching etcetera that is there was the part of the cloud when you want to use those services which we have not coded it has been developed by the cloud service provider. How will we do this indexing or say searching on an encrypted data base by using a non proprietary that is a generic software provided by the cloud? The cloud when we hire a cloud they gives certain services it may include this index and search.

So, in this context there are there are lot of research that is happening. We will talk about these encryption algorithm down the line in our phase 4 or phase 5 courses. So, what does these encryption algorithms two there is something called an homomorphism encryption. What will the homomorphism encryption do? It will encrypt your data into cyber text, but the data can be analyzed and you can also do some operations on that data, like your indexing searching as if you can do it on the normal data. So, this is an interesting. So, I need not decrypt to actually do certain operations. I can use the same encrypted data to do certain operations and basically get the result. And of course, when we display the result there can be; post that there could be wrapper that can be decrypt and give you there actual data. So, by having such a homomorphism encryption what do we achieve?

Some generic software which is written which does not have any knowledge about our encryption can be used seamlessly on our encrypted data. And this is very very relevant when we look at cloud. The next thing is called predicate encryption. Predicate is nothing but is secret function f . So, I would like to know whether a particular data satisfies some property. So, I have a secret function f to which I give the encrypted data as an input. And if that function evaluates to 1 that means, I can say it satisfies that particular data item encrypted data item satisfies the property. If it evaluates to 0 then it does not, but that function f will be known, it is a secret key function.

So, by this what could happen is that I could go and these type of predicate encryption is something which is generic. I need not do a decryption. I can just go and evaluate some function and then when it comes out 1 or 0 based on that I will know something about the data some characteristics about the data, which is sufficient for me to do for that processing. So, this homomorphism encryption and a predicate encryption are two interesting encryption research avenues today which can basically help you, encrypted data and put it in a common infrastructure and do some operations on that encryption n data encrypted data without decrypting. So, this is very very important. So, and then there is a need for when you process the data especially in a multi tenant model, there is

a need for a security. So, so these are all some aspects when we start looking at data even at the generation phase as it moves from the client to the cloud. Now, these are that data can be in transit data can be in rest and at all stages we need to go and have an encrypted form. And sometimes without decrypting it we can basically go and do lot more operations when we use some specific encryption algorithms. And lot more security need to be taken when especially while processing data and that too in a multi-tenant module. Now, the other thing is that somebody who does not there should be a sort of accountability for this data to talk about integrity.

(Refer Slide Time: 08:41)

Data Security and Storage (cont.)

- Data lineage
 - Knowing when and where the data was located in the cloud is important for audit/compliance purposes
 - e.g., Amazon AWS
 - Store <d1, t1, ex1.s3.amazonaws.com>
 - Process <d2, t2, ec2.compute2.amazonaws.com>
 - Restore <d3, t3, ex2.s3.amazonaws.com>
- Data provenance
 - Computational accuracy (as well as data integrity)
 - E.g., financial calculation: $\text{sum} \left(\frac{((2*3)*4)}{6} - 2 \right) = \2.00 ?
 - Correct : assuming US dollar
 - How about dollars of different countries?
 - Correct exchange rate?

Thought bubble text:
 Where is the data that system located?
 What was the state of that physical system?
 How would a customer or auditor verify that data?

So, the two terms are introduced from the cloud point of view what you call as data lineage and data provenance. What is data lineage? The data lineage essentially means that you would like to know when and where the data was located in the cloud right. So, for example, the amazon web services AWS use you for data something like store the day time and from where and process restore. So, for the exact lineage from where it originated where and all it stayed and where it came that should be available as an audit. So, that at some point of time when I want to go and ensure integrity of the data and these type of logs are very very important.

And also the next thing is data provenance. Here we have to go and look at the computational accuracy of the data. For example if we are doing a financial calculation like

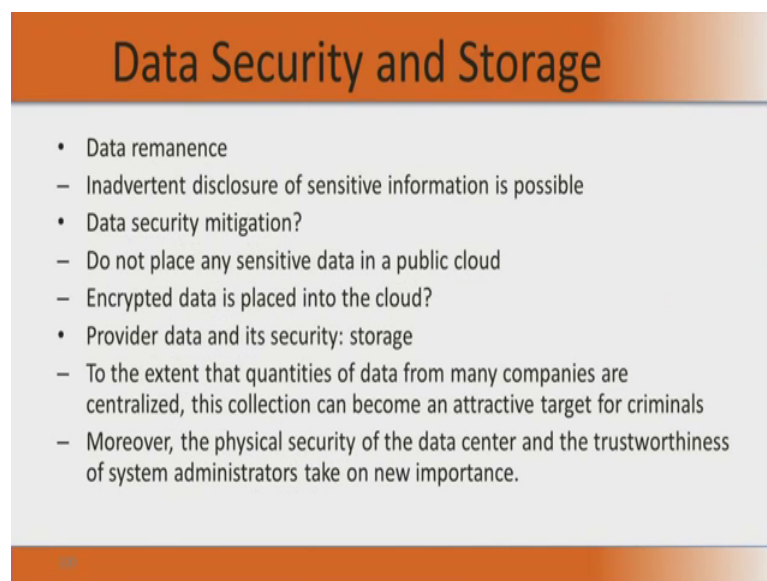
$$\left(\frac{(2 \times 3 \times 4)}{6} \right) - 2 = \$2 \quad (9:51)$$

Now, the simple thing you have to notice the 2 into 3 into 4 into 6 is in what currency and the other 2 minus 2 is in what currency both should be in same currency.

So, if I just say dollar one can be Singapore dollar can another can be say American dollar. So, I have get back some correct exchange rate. So, these type of computational accuracy need to be also certified when we try and use third party software like cloud service provides some software. Then there is no way by which you can test. So, these things or something that we need to keep in mind. Otherwise this can spoil the integrity of your data.

So, when we move from a normal data center to a cloud type of infrastructure these are this is one very important thing. So, as you see there in your thought right there is where is the system located what was the state of that physical system, how would a customer or auditor verify these information about systems and their is physical state. This is what is a challenge that these to be addressed and many of this are taken care of by having logs and also having some very carefully coded procedures.

(Refer Slide Time: 11:21)



The slide features a title bar at the top with a gradient from orange to white, containing the text "Data Security and Storage". Below the title bar is a light gray rectangular area containing a bulleted list. The list items are: "Data remanence", "Inadvertent disclosure of sensitive information is possible", "Data security mitigation?", "Do not place any sensitive data in a public cloud", "Encrypted data is placed into the cloud?", "Provider data and its security: storage", "To the extent that quantities of data from many companies are centralized, this collection can become an attractive target for criminals", and "Moreover, the physical security of the data center and the trustworthiness of system administrators take on new importance." At the bottom of the slide, there is a small orange bar with the number "10" in white.

Data Security and Storage

- Data remanence
 - Inadvertent disclosure of sensitive information is possible
- Data security mitigation?
 - Do not place any sensitive data in a public cloud
 - Encrypted data is placed into the cloud?
- Provider data and its security: storage
 - To the extent that quantities of data from many companies are centralized, this collection can become an attractive target for criminals
 - Moreover, the physical security of the data center and the trustworthiness of system administrators take on new importance.

10

The other important issues when specifically while we are storing data, we go and erase, but is it getting completely erased actually there is a property call data remanences. Remanence which is that remaining after that the data remaining still after you go and erase some magnetic field, it does not immediately get erase and it will remain for some time. And if I have this data remanence then some inadvertent disclosure of sensitive information is possible.

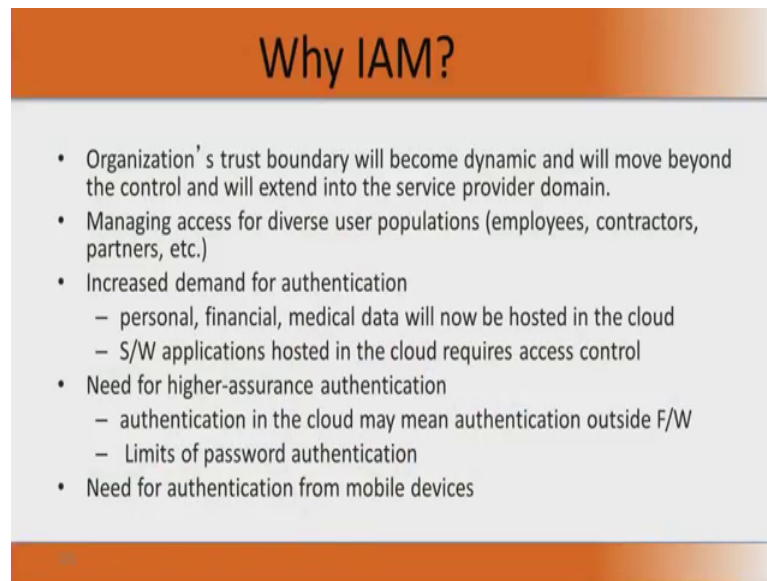
Though I think I have deleted it it is not actually deleted it still some remanence of those data is still there. And you basically can get some sensitive information. So, the other thing is that when we move from a normal cloud normal data center to a cloud, all your data all your file system since they are going to be on a public place in a cloud need to be encrypted and then decrypted, if necessary. Otherwise some homomorphism encryption or the or the predicate type of encryptions are used.

We need to work on encrypted data to do the computations right. So, the other thing is that to the extend quantities of data from many companies are centralized right. This collection can become an attractive target for criminals. So, a cloud is something very interesting for a criminal because if we gets into that is not getting into one company, but we can get into many companies data. So, the other thing is of course, you are talking about people, process, technology.

People the physical security of the data center all the and. So, data centers are rated as tier 1 tier 2 tier 3 extra depending upon these features available. So, these are all standard documents. So, there is a tier 1 data center, tier 2 data center etcetera. And what you mean by this? So, their levels of security are higher and then the people there the trustworthiness of system administrators and takes very very good and the physical security at the data center are prime importance in this effort.

When there is a movement from a normal data center to a cloud the most important thing that comes up other than the other sec the other security we have taken care of is the identity and access mechanism.

(Refer Slide Time: 14:11)



Why IAM?

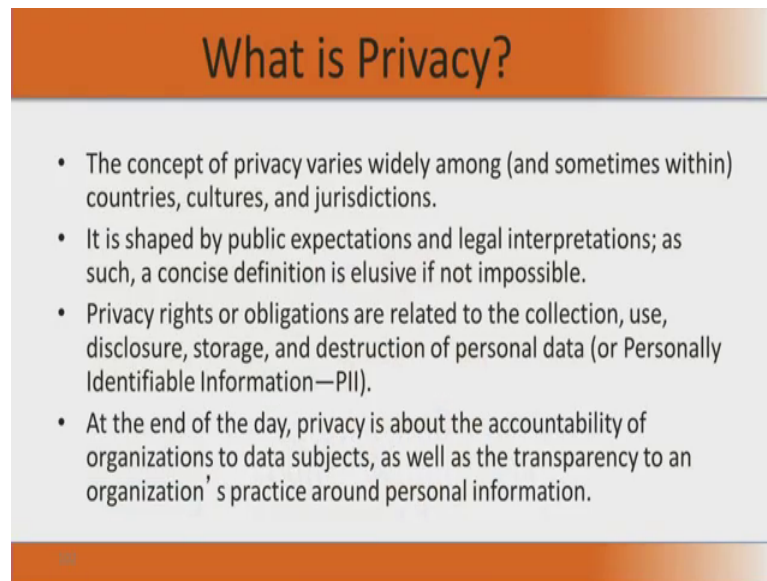
- Organization's trust boundary will become dynamic and will move beyond the control and will extend into the service provider domain.
- Managing access for diverse user populations (employees, contractors, partners, etc.)
- Increased demand for authentication
 - personal, financial, medical data will now be hosted in the cloud
 - S/W applications hosted in the cloud requires access control
- Need for higher-assurance authentication
 - authentication in the cloud may mean authentication outside F/W
 - Limits of password authentication
- Need for authentication from mobile devices

What happens when we move from a data center to a cloud which is more public? Actually your trust boundary is now enlarged. We have still now depending upon trusting the system integrator. Suppose you have a data center there is a system integrator for you, there is a network service provider for you. We were locate we were just believing or trusting these people now the trust is now expanded to the cloud service provider.

So, we go into his domain and from the cloud service provider point of view he has many clients. Each client has many employees contractors partners etcetera. And so, he has to now go and manage the access of this diverse set of population across different customers. And that also in that context also when I want to go and access a cloud the identity and the related access management. So, I identify a user, I authorize in, he authenticates and I once he is authenticated how do you grant access. So, this IAM becomes very very important as we from a cloud service pointer or cloud service provider point of view.

So, what happens is because of many many people trying to access this cloud there is a need for doing lot of authentication which is more in a personal. So, there is an increase demand for authentication. And there is also a need for high assurance authentication. It is should not be just password, but in a multiple levels of authentication. And there should be special types of authentication for mobile devices how to handle? These are all issues that come.

(Refer Slide Time: 16:06)



What is Privacy?

- The concept of privacy varies widely among (and sometimes within) countries, cultures, and jurisdictions.
- It is shaped by public expectations and legal interpretations; as such, a concise definition is elusive if not impossible.
- Privacy rights or obligations are related to the collection, use, disclosure, storage, and destruction of personal data (or Personally Identifiable Information—PII).
- At the end of the day, privacy is about the accountability of organizations to data subjects, as well as the transparency to an organization's practice around personal information.

Now, let us go and look at the very very important thing of privacy. We have not talked about protecting data, but one important aspect of protecting data is to protect the privacy that the data essentially gives. Now, one of the bigger challenges like privacy trust. Trust and privacy are not the same they are different. And as we go down we will see much more concise definition for privacy, but now as we said there is a definition of trust is always elusive in in the previous session.

We did mention that trust does not have a mathematical formulation similarly privacy also a sort of a very a evasive definition is what we can give because privacy is still not well defined. The concept of privacy will vary among countries what is private to one country may not be considered as a private information in the another country. And what is it what do we meant by a privacy? I go and assure privacy means privacy is for the entire data cycle I collect the data I use the data I disclose the data I store the data I destruct the data. And this data if it is personal then there is something that we have to do at every stage.

So, that gives as to give a very a short a very constructive name for this we call this data as a personally identifiable information. So, a PII. So, we have to be very careful about handling PII. So, finally, if you note all this privacy is about the accountability of organizations to that data subjects. As well as the transparency to an organizations practice around personal information. How do we handle personal information?

So, the concise definition of privacy is very difficult, but these are all the notions of

privacy that we are talking of.

(Refer Slide Time: 18:25)

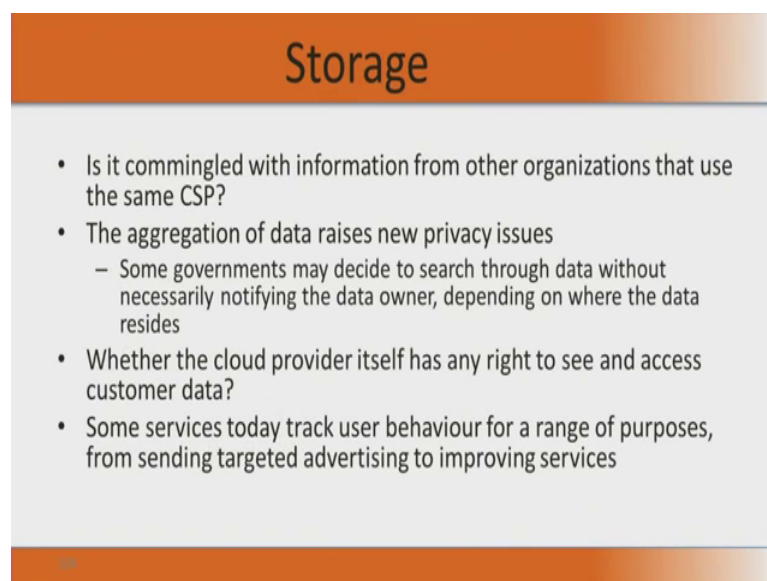


What Are the Key Privacy Concerns?

- Typically mix security and privacy
- Some considerations to be aware of:
 - Storage
 - Retention
 - Destruction
 - Auditing, monitoring and risk management
 - Privacy breaches
 - Who is responsible for protecting privacy?

Now, we do not we should not mix security and privacy. Privacy is something much more different. We will understand what are what are the difference between privacy and security by some of the needs that privacy enforces us. For example, there is a need for privacy at the storage stage, at a retention stage, at a destruction stage, at an auditing monitoring and risk management stage. And then we will look at some breaches privacy breaches and who is responsible for protecting privacy. We will look at all these things as a part of understanding privacy.

(Refer Slide Time: 19:12)



Storage

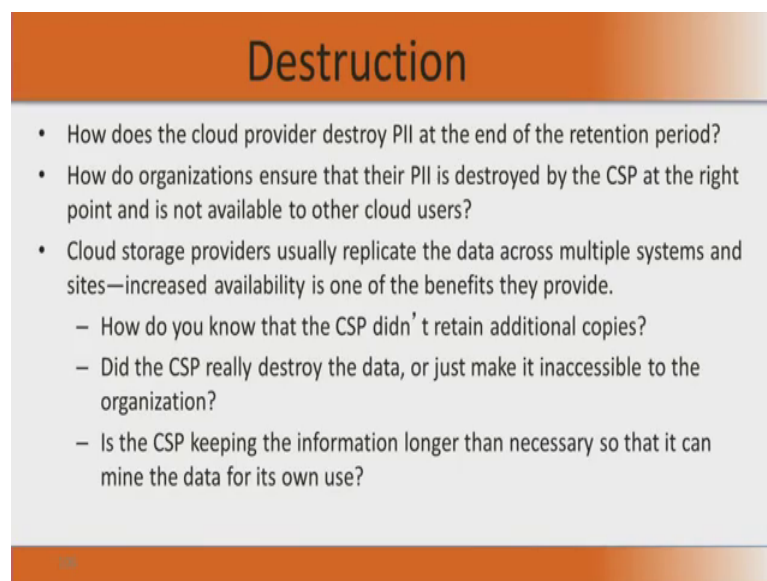
- Is it commingled with information from other organizations that use the same CSP?
- The aggregation of data raises new privacy issues
 - Some governments may decide to search through data without necessarily notifying the data owner, depending on where the data resides
- Whether the cloud provider itself has any right to see and access customer data?
- Some services today track user behaviour for a range of purposes, from sending targeted advertising to improving services

So, we will look at what is privacy from a storage angle. Privacy from a storage angle we have already talked of in great detail. And today when we look at these type of multi tenant model I am storing data on the same disk. I am storing data on the same memory, but now there is some virtual machine which is going to provide be this isolation and will it provide be the isolation is the first consideration. So, the aggregation of data rises new privacy issues because there will be many many customers would be sharing a particular platform and the same hardware physical platform.

That is going to be a major privacy issue in storage and. So, somebody says that my privacy of data is ensured then they should really go and prove, that there is indeed an isolation in different programs or different virtual machines, storing data on a storing their data on a single physical disk. So, this is something that needs to be there. And talking about retention how long is personal information retained? What is the retention policy?

Does the organization own the data or the cloud service provider once the data who enforces the retention policy in the cloud and how are exceptions to this policies managed. Such as if there is a litigation I cannot go and destroy the data. So, how are these things actually managed? So, this is another thing that we need to take into consideration when we are looking at privacy and destruction.

(Refer Slide Time: 20:56)



Destruction

- How does the cloud provider destroy PII at the end of the retention period?
- How do organizations ensure that their PII is destroyed by the CSP at the right point and is not available to other cloud users?
- Cloud storage providers usually replicate the data across multiple systems and sites—increased availability is one of the benefits they provide.
 - How do you know that the CSP didn't retain additional copies?
 - Did the CSP really destroy the data, or just make it inaccessible to the organization?
 - Is the CSP keeping the information longer than necessary so that it can mine the data for its own use?

So, how does the cloud provider destroy? When you destroy have you destroyed everything are they key keeps a ball part a keeps a copy and then destroy? And if he

keeps a copy what is be there implications? So, these are all something that we need to do the privacy at that destruction stage.

(Refer Slide Time: 21:18)



Auditing, monitoring and risk management

- How can organizations monitor their CSP and provide assurance to relevant stakeholders that privacy requirements are met when their PII is in the cloud?
- Are they regularly audited?
- What happens in the event of an incident?
- If business-critical processes are migrated to a cloud computing model, internal security processes need to evolve to allow multiple cloud providers to participate in those processes, as needed.
 - These include processes such as security monitoring, auditing, forensics, incident response, and business continuity

And the more important thing is about auditing, monitoring and risk management. So, how as an organization, suppose let me say there is a cloud service provider I am organization A. How will A now go and audit X? When it starts auditing X, has X does not have can I go and auditor audit only my virtual machine or can I go and audit only my data. So, that is another thing and do you actually audit it regularly and what happens when. So, how do you go and audit first and if at all you go and audit how do you audit it regularly. And if there is an event what do what do what happens if there is an incident here right.

Specifically when business critical process is are migrated to a cloud computing model, the internal security process need to be evolved to allow multiple cloud providers to participation those processes. So, these include processes such as security monitoring, auditing, forensics, incident response and business continuity. All these things again we have to go and look at from a cloud point of view. Now, this is also privacy concern because the moment I am going to audit or monitor there is a notion that I am trying to sneak into data. So, how is it that in spite of this auditing and monitoring your data is still, your privacy is not breached.

So, in the next session we will talk more about privacy breaches.