CHAPTER 2

THEORITICAL FOUNDATION

2.1 Cloud Computing Architecture

2.1.1 Brief History

The idea of cloud computing was born from the need of sharing data for the people around the world. According to Mohamed [7], J.C.R Licklider as the initiator, wanted to enable people to access anything from anywhere. This is similar to what we know as cloud computing today. Another figure who contributed to the history of cloud computing is John McCarthy. He was the pioneer of artificial intelligence and mathematical theory of computation. He proposed the idea of utility computing. He wanted that computing can be used as a public utility. And with the birth of grid computing, the cloud computing via the Internet became reality.

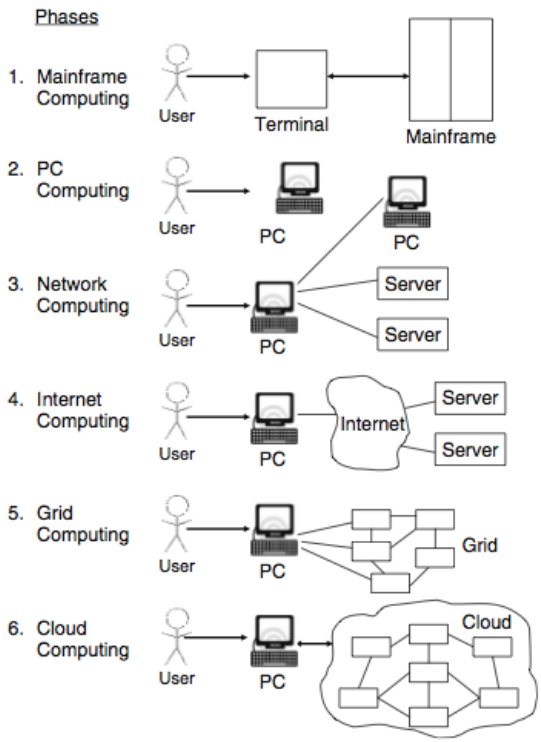

Fig. 1.1 Six computing paradigms – from mainframe computing to Internet computing, to grid computing and cloud computing (adapted from Voas and Zhang (2009))

Figure 1. Six Computing Paradigms [8]

The picture, which is taken from the Handbook of Cloud Computing by Furht [8], shows that the computing paradigms change over time. At first, used the mainframe computer to perform computation. They accessed the mainframe computer through a terminal. Then the giant mainframe computer was replaced by personal computers. Users can have their own personal computer to connect to other personal computers. And next the server is invented to store the data from the personal computers. In other words, the computers are connected using a server. When the Internet begins to boom in the world of computing, the Internet is used as the connector between the servers, which eventually connect all the personal computers.

Cafaro [9] claimed that Grid computing is a method of computation where several machines act as a pool for resource sharing and ultimately as the problem-solving infrastructure. Using the grid computing, the computers are connected together via the Internet and able to do single processing. And from the grid computing, the utility computing is developed. It is a technology where people can get what they want from anywhere. This will later become the foundation of the cloud computing. And the last piece of the puzzle would be the virtualization. Virtualization is considered as one of the key functions in cloud computing.

2.1.2 Cloud Delivery Models

In line with Furht [8], there are three cloud delivery models which differentiate the services and level of risk maintained by both the providers and the customers. They are the SaaS, PaaS and IaaS.

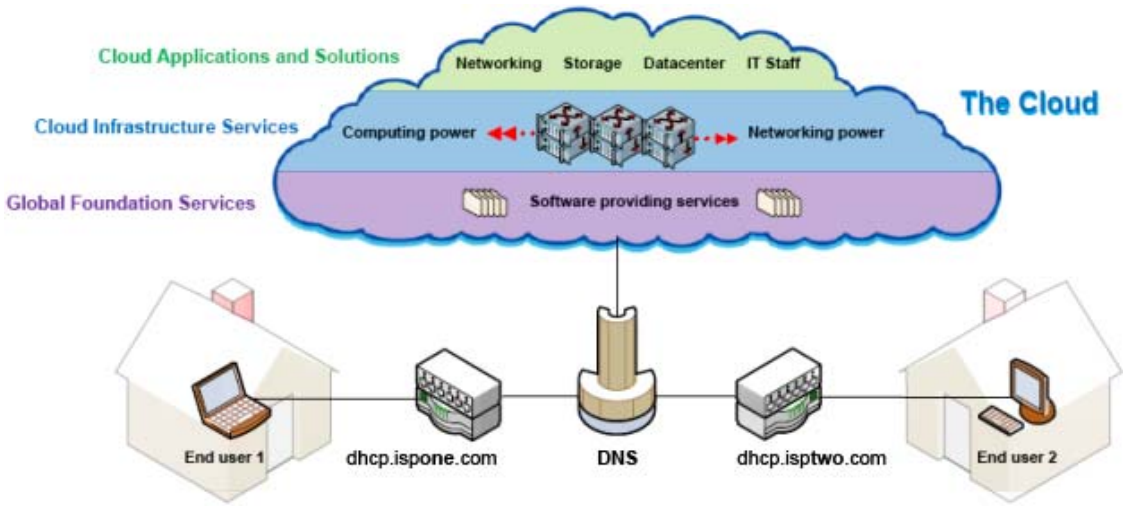

Figure 2. Generic Cloud Computing Supporting 2 Users [10]

Software-as-a-Service (Saas).

Using this model, the customers are put at ease by the cloud provider. The cloud provider prepares the applications and the infrastructures for the customer. The example of this model is GoogleApps and Salesforce.com. The main benefit of this model is building the platform and infrastructure of the cloud less troubles the customers. They only need to submit their content of the site to the provider. However, the customers must follow the procedures made by the provider. As an example, the customers who want to use GoogleApps service must write the code using Python language.

Platform-as-a-Service (PaaS).

Using this model, the customers can build and develop their own application and put it in the cloud. In other words, the cloud acts as a virtual development environment for their applications. This model enables better control over data, application and security. One example of this is the Amazon Web Service (AWS).

Infrastructure-as-a-Service (IaaS).

Using this model, the customers will have more control over their applications. The cloud provider provides processing, storage, networks and other computation essentials to the customers. It also enables the customers to install and maintain their own operating systems and applications. But the customers do not have the control over the infrastructure of the cloud. They just lease the hardware and software services from providers such as Rackspace or GoGrid.

2.1.3 Cloud Deployment Models

In keeping with Furht [8], cloud also has several deployment models. These models distinguish the ownership of the cloud network itself, whether it is public, private, community or hybrid.

Public cloud

Available to anyone and can be accessed via the Internet. This type of deployment model is more like free-for-all model that can be used and accessed by all people. But the public cloud is more than just a cloud for everyone. Some companies also place their private cloud in the public cloud. So instead of setting up new private cloud, those companies can just register and use the service of other cloud provider to put their private cloud in the network.

Private cloud

Created and developed by single organization for own purposes. Usually a company owns a private company for the purpose of ease of access. So the people within the company can access the data at the company's private network from anywhere. And this private cloud can also be included into a public cloud.

Community cloud

Created and developed by several organizations that become a community and shares the same purpose. The concept of this community cloud is similar to private cloud,

where only registered users can access the cloud. But the difference is this type of cloud is shared among several companies.

Hybrid cloud

This model is the combination of two or more deployment models. The most common combination used is the combination of private cloud (for managing the internal cloud) and the community cloud (for sharing among entities). It means that within the community cloud, there can also reside the private cloud of a company, which is also the member of that community cloud.

2.1.4 Cloud Computing Characteristics

Grobauer et al [11] define the unique characteristics of cloud computing. These characteristics are also related to the vulnerabilities in the cloud itself. Those characteristics are the on – demand self service, ubiquitous network access, resource pooling, rapid elasticity and measured service.

On – demand self service

The users will not have any trouble whenever they need something from the provider. Without contacting the provider directly, the user can still request for service. The users can request the service through web portal, management interface, etc. The provider will handle the requests later. In other words, cloud becomes the bridge between the provider and the users with less human interaction.

Ubiquitous network access

Since cloud is connected to the Internet and based on the Internet, all the services are available on the Internet. And since Internet is available almost everywhere with the development of technology, cloud becomes available almost anywhere. It also does not require complex protocols or mechanisms.

Resource pooling

The infrastructure in the cloud provides the resources that can be retrieved by the users. Given that the infrastructure is homogenous and can be shared among all users, the resources are also shareable among all users. With the combined and shareable resources, this can improve the performance of the cloud.

Rapid elasticity

The most powerful benefit of cloud computing is the scalability. It means that the resources in the cloud can be scaled up or down easily according to the need of the user. To increase or decrease the capacity of the cloud does not require complicated process, since the resources are shareable among the users. This can improve the efficiency and effectiveness of the resources in the cloud.

Measured service

The resources used in the cloud are always measured. Using the metering or billing method, every usage of the resources by the users can be detected. So the resource

optimization can be achieved and the usage report can be delivered accurately to the customers.

2.2 Principles of Infrastructure Security

King [12] stated about the importance of infrastructure security. This covers the security within the environment of a system. And this infrastructure security has several components, which will be explained further below.

2.2.1 Infrastructure Security

First of all, the basic understanding about the infrastructure security must be understood. The security within the infrastructure is not the same as the security in a component. The security mechanisms in the infrastructure will work only if they are connected to other aspects of security in the infrastructure. Unless they are connected and interact with others, they are called a security component, not a part of security infrastructure.

2.2.2 Components of Infrastructure Security

There are four main categories of the infrastructure components:

- Network
- Platform
- Physical
- Process

The network consists of the firewalls, switches, routers, and other security devices that increase the security level of the overall infrastructure. These components are used to manage the traffic in the network. The purpose of this network level devices is to help monitoring and maintaining the security of data, application and network within the infrastructure.

The platform mainly consists of the server and client-side software. One example of this category is the Smart cards and the readers, where the Smart card is used as the authentication and authorization device and the readers as the system that will grant the privileges to the cards. The main purpose of this platform level security is to provide the authentication – authorization and protect the major infrastructure boundaries.

The physical components of the infrastructure security consist of the door locks, security cards, CCTV and other physical security peripherals such as the biometric components. The emergency and backup systems such as the UPS are also part of this category. The purpose of this physical level is to avoid unauthorized entities to enter the security perimeter of the infrastructure and to maintain the availability of the system in case of electricity failures.

The last component is the process components of the infrastructure. This category mainly consists of security policies and procedural documentations that govern the systems and networks where the corporate data resides. The purpose of this process level is to provide the scope of protection and act as a guide for the whole organization. This

is essential to prevent the social engineering attacks on the employees. More importantly, these policies will be useless without the proper understanding of security infrastructure goals.

2.3 Virtualization

2.3.1 Definition of Virtualization

According to Waters [13] virtualization is defined as,

"*Technologies designed to provide a layer of abstraction between a computer hardware systems and the software running on them, by providing a logical view of computing resources rather than a physical view*"

To put it simply, virtualization hides all the physical information of the system from the users. They can still interact with the operating system, memory and other resources without interacting with the physical devices directly. There are many things that can be virtualized. Memory, server and even a desktop can be virtualized.

2.3.2 Benefits of Virtualization

Perilli [14] listed out several benefits of virtualization for cloud computing. Basically there are many things that virtualization has to offer for business. Virtualization eliminates the traditional computing paradigm that is one server will have one processor and one operating system. But using virtualization, a server that has the processor of multiple cores, can have multiple operating systems running on them. Furthermore, it can be leveraged to suit the business.

Using virtualization, older software or legacy software can still be used and run on newer hardware. This could be important since some companies might find the legacy software indispensable from their current business model. Also, the requirements for hardware, power and storage can be adjusted to produce lower operating cost.

As for the security, virtualization can ensure better security compared to the traditional computing. By virtualizing, the workloads can be distributed quickly to increase the fault tolerances. Concurrent access can also be handled better using virtualization. In addition, it also helps in the disaster recovery process by providing redundancies. It is now clear that virtualization also provides some security benefits to ensure the business competency and business continuity.

2.3.3 Virtualization Approaches

Scheffy [15] pointed out that all the approaches require the *hypervisor* software, which is used to allocate the basic machine resources including CPU time and memory.

Full Virtualization

The full virtualization means that the hypervisor is used to capture the machine operations used by the OS to do activities on the system. These activities can be reading operations, modifying operations and even the input – output operations. After the machine has been captured, the hypervisor will emulate the process so it resembles the real hardware. The advantage of this approach is the capturing and emulation process is hidden from the users. They will not notice the changes made in the system. In addition,

it can be installed on any guest OS. However, it also comes with a disadvantage, which is the slow performance. The capturing and emulation process can slow down the performance of the system. These processes will create a layer to perform each capturing and emulation. Thus it can reduce the performance depending on the workload.

Paravirtualization

This approach is also called partial – virtualization. It eliminates the capturing and elimination process performed in the full virtualization approach. Instead, it will make the guest OS to cooperate and create the virtualizing illusion. The advantage of this approach is in terms of performance. It can perform faster compared to other approaches. However, it requires a uniquely modified guest OS to be able to create and maintain the virtualizing illusion.

Hardware – Assisted Virtualization

As the name suggest, this approach relies much on the hardware extension on the architecture. This hardware extension will eliminate the need of the hypervisor to do the capturing and emulation of the operations. One example of this approach is by using AMD Virtualization (AMD-V) feature from AMD. This feature will enable the virtualization using the hardware assistance. For some operations, the hypervisor will still need to do some emulation. But it can be handled by creating the virtual mapping of the input and output devices. However, it does require some modifications on the chipset.

2.3.4 The Role of Virtualization in Cloud

Vaquero et al [16] stated that the cloud is a massive space that contains a lot of virtualized resources (such as hardware, development platform and/or services). These resources are available and can be adjusted to meet the business requirements optimize the resource utilization. The resources are brought using the pay-per-use model, which is guaranteed by the Infrastructure Provider in the form of customized SLA (Service Level Agreement). According to this definition, the virtualization provides the dynamic environments and SLAs are very crucial for the cloud computing itself. This also drive a new business model where you can have what you need (on-demand service) by just leasing what you need (pay-per-use) without having to completely build the infrastructure.

As virtualization plays an important in cloud computing, there are specific roles that the virtualization provides in cloud computing. The first and foremost, it hides the infrastructure and other information from the customer, which is the signature of cloud computing. By making everything virtual, the customers will only see the result. They will not know about the infrastructure at the data center. They will have the SLA to ensure that their money will not go into waste. Cloud and virtualization will guarantee the users for the infrastructure.

And then the virtualization increases flexibility, elasticity and scalability to optimize the efficiency. Most people know about this already since it becomes the unique features that the cloud has to offer. The users can easily scale the infrastructure to suit their

business needs without any trouble of server downtime or maintenance. In other words, there is no significant disturbance to their business in cloud. They just have to request it and the providers will do it for them in a quick manner. This feature can eliminate the problems of expansion by increasing efficiency.

In the previous point, the virtualization can help to drive the business. Specifically, the virtualization can help to develop new business strategies. The users can add a new business model in their existing services or even change the business model into a more sophisticated way. Obviously, this can expand the business competency.

Using virtualization, the Operating System layer and the application layer is separated from the hardware. This is related to the scalability where the servers can be easily added without any trouble. By separating the hardware and software, those two are not strictly bound to each other since all of them are virtualized. One hardware can support more than one software, which means more hardware deployed will result in much more software running at a time. So it is not one Operating System for one server only. This is essential to support the on-demand services.

Another role of the virtualization is to provide flexibility. Although many people stated that it is for cost saving, actually it is not. It is true that virtualization can reduce their cost of setting up the server and maintenance. But more than that, the virtualization can eliminate the cost for expansion. It means the users will not be troubled with the

expansion. As stated in the points above, they can just contact the provider and they will do it for the customers. It does save some cost, but it is more into flexibility.

2.4 Threats in Organization

These are the threats that usually disturb the security in an organization according to Whitman [6].

2.4.1 Acts of human error or failure

Human is the weakest link in the system. Accidentally or not, the threat of human error is usually high. For example when someone accidentally deletes an important file that can impact the business greatly. Or someone forgot to log out after accessing the central database, making it vulnerable to attacks. There are many ways for attacks or disaster to occur through this threat. What the company can do is to give education and increase the level of awareness of the employees. At least that can reduce the chance of human error.

2.4.2 Compromises to intellectual property

Intellectual property is the result of someone's creativity and can be subject to copyright or patent. Today, breaking the copyright rules is very common. There is a group of hacker called *cracker* who aims to crack the license of software. This attempt to compromise the intellectual property is considered as threats, since cracked software can be attached with malware. And when viruses infect the customers after using one company's software, the business repute of that company can be damaged. Or the cracker can illegally duplicate the software and sell it with cheaper price. Obviously, the

customers might prefer to buy the software at cheaper price. This can be dangerous for a business to lose their customers. A good software security must be implemented to avoid this threat.

2.4.3 Deliberate act of trespass

Trespassing is the act of gaining access to the system illegally. When an outsider gained access to the system, that person can do many harmful things to the system. The hacker can install a Trojan, viruses, worms, etc. to damage the system. Furthermore, the hacker or cracker can also collect the confidential data of that company. Beside those two, there is also another type of hacker called *phreaker*. This type of hacker will gain access to the phone lines of the company and use them to make phone calls illegally. The company can suffer a vast amount of phone bills without knowing about it. A proper configuration of firewall, routers and other security mechanisms can help to reduce this threat. For example the firewall will help to filter the traffic coming in and out of the network. Hopefully, unauthorized person will have difficulties to gain access to the system.

2.4.4 Deliberate act of information extortion

Information extortion is when someone threatens a company or an organization using blackmail. The hacker will demand something in exchange for the information the hacker has successfully retrieved illegally. The demand is usually in the form of money and the information is usually top classified such as critical data or credit card numbers. The company will be forced to follow the demand and pay the money. If not, the information can be leaked further and become a threat to a person or a company. Even if

they have fulfilled the demand, nobody can ensure that the information will not be leaked. The source of this threat can be internally or externally. Nonetheless, both sources can endanger the company. So the best thing to do is to prevent this to happen.

2.4.5 Deliberate act of sabotage or vandalism

Sabotage or vandalism is the destruction of an asset or image of the company. The destroyed assets can be servers or network devices. This can endanger the availability of the system. A physical security by placing security and locks can reduce the chance of this threat to occur. The other form of vandalism, which is the image destruction, is considered more dangerous for the business. It can cause customer distrust to the company and eventually make the company to lose their customers. The nastiest form of this threat is cyber terrorism, just like the one happened to the U.S. government.

2.4.6 Deliberate act of theft

Act of theft means the act of taking someone's property illegally. Similar to sabotage or vandalism, this threat can occur in physical form and digital form. Protecting the assets physically is a must and an investigation procedure after an incident happened must be well planned. However, protecting digital assets can be more difficult. Beside the system must be equipped with security a mechanism, the investigation is more difficult to be carried out in digital form. The digital forensics is the only way to trace and investigate the digital theft.

2.4.7 Deliberate software attacks

The software used to attack or do harmful things is called malware. The malware can be in the form of Trojan, worms, virus, back door and many more. The intent of this malware is to disrupt the business in terms of lost of availability, such as the Denial of Service attack. If the customers cannot access the system, it will obviously cause a riot in the customers and lead to customer distrust. In addition, the DoS attack is usually hard to trace. So it can be difficult for the company to find the mastermind behind the attack. A proper security must be implemented to prevent any authorized access. Unauthorized access will help the hacker to plan the attack better and launch a devastating attack.

2.4.8 Forces of nature

The natural disasters can also become a threat for the business. If the building or system is damaged because of flood or fire, the business can be damaged as well. Since it is quite difficult to predict the nature, the company can only prepare the contingency planning if this incident occurred. They must ensure that the business can still run when attacked by natural disaster.

2.4.9 Deviation in quality of service

In this case, the deviation means that the product is not performing as expected and disrupts the business process. For example, the online banking system is disrupted because the Internet connection of the company is very slow. And when a lot of customers are accessing the system, the system cannot handle the load and start to cause

errors. This can be crucial for online businesses since it relies a lot on technology. A proper quality control and quality assurance might help to minimize this threat.

2.4.10 Technical hardware and software failures

The hardware and software failures can occur when the company bought hardware or software that contains fault or flaws. When the hardware and software is used in the day-to-day activity, they do not perform as expected. Moreover, the combination of several hardware and software might reveal another bug. It is the best for the company to check the compatibility of all hardware and software. Not to mention the quality control and quality assurance of those hardware and software.

2.4.11 Technological obsolescence

Older hardware or software sometimes cannot cope with the advanced environment or newer technologies. And this obsolescence can cause malfunctions or errors. However, some companies still use older technologies because the technology they need is there. Or maybe the employees will have difficulties to adjust and adapt to newer technologies. So the presence of legacy software is considered important. A proper managerial planning must be done to maintain the technology at its peak performance since IT plays a large role in the business.

2.5 Risk Control Strategies

Whitman [6] identifies that there are four risk control strategies that must be chosen in order to handle the risk. Those four strategies are avoidance, transference, mitigation and acceptance.

2.5.1 Avoidance

Risk avoidance is the strategy to avoid risk by prevents further exploitation to vulnerabilities. So it will try to remove the risk that threatens the known vulnerabilities. Obviously, in order to be able to avoid the risk, the vulnerabilities of the system must be known. Else it will be the difficult to avoid unknown risk. It is also the common preferred approach to control the risk. The strategies to avoid the risk can be by applying policy, educating the people or applying technology.

Applying policy means creating the rules and regulations that must be known and understood by the people. For example, there is a rule that the employees within a company cannot open certain websites during office hour. Or the employees can only download a file with size smaller than 10 megabytes for example. These policies are meant to control the traffic in the network of that company. It will certainly help to avoid unwanted risk.

Educating the people is also very important. Generally, only the technicians or IT people who know about the threats and vulnerabilities. In fact, the company consists of people from other department without IT background, such as the accounting, marketing or

human resource. Those people without the IT background must be educated so that they are aware of those threats and vulnerabilities. This is also important to reduce the probability of internal threats such as human error, which can be very dangerous.

Another one is to apply the technology. It can be firewall, IDPS, routers and many more. Those technologies are used to manage the traffic actively and usually used as the frontline in the security of the infrastructure. The cooperation of those devices will help to cover the vulnerabilities of the system or to avoid the threats.

2.5.2 Transference

Transference is the method of controlling risk by allocating the risk to be handled by third parties. Usually this step is taken if the risk is too difficult to handle or the risk is already affiliated with the third party. An example of this would be the server maintenance. If a company sets up a server on its own, they will have to do the maintenance on their own as well. But if they hire the server hosting service, the problems and maintenance will be taken care by the hosting company. Not only the maintenance, but also the other risk associated with it. However, this control strategy requires the company to trust the third party to handle the risk. Or else, it will just become another risk.

2.5.3 Mitigation

Mitigation is the strategy to reduce the impact of the risk, particularly caused by vulnerability exploitation. The focus of this strategy is to reduce the damage so the

business can still run regardless of any occasion. The risk mitigation consists of three plans and they are related to each other; Incident Response Plan, Disaster Recovery Plan and Business Continuity Plan. All those plans rely on the ability to detect and respond to attack quickly as well as the coordination with other plans. Below is the explanation of each risk mitigation plans.

The Incident Response Plan (IRP) is the list of plans that must be taken should a disaster happened. This will give the clue on what to do when a disaster or incident happened. For example, when there is a fire in the building, they already know where to go. Another example is backing up the data as quick as possible whenever the system got compromised or attacked. They know how to act and react according to the IRP. It will also lighten the workload of the rescue team, knowing that the people will cooperate with the evacuation. The timeframe of this plan is immediate and more to real – time action.

The Disaster Recovery Plan (DRP) gives details of the plan on recovering after the incident occurs. Usually it takes place after the incident is finished and things are starting to get back to normal. For example, the first thing to do for the banking company is to restore the data from the backups or planning to renovate the building after the incident. The focus of this DRP plan is to get back on track as soon as possible and avoid further loss in short term.

The Business Continuity Plan (BCP) encompasses the plan to keep running the main business process while the incident takes place. It is also the steps taken when the scale of disaster exceeds the capability of the DRP plan. So the customers will not be worried about the downtime caused by an incident. For example running the servers or business processes on emergency sites. Even though those emergency sites might not be able to support the whole business function, at least the company can still run the business at the state where the customers will not be worried. The focus of this BCP plan is to avoid disruption to the core business process of the company. Whitman [d] added that this is the most strategic and took the longest time among the three risk mitigation plans.

2.5.4 Acceptance

The risk acceptance is probably the most arguable risk control strategy. The perception of risk of many people might be different. Some might consider a risk will not have a major impact, but for others that risk might be dangerous. It will depend on the risk appetite of a company to accept that risk. Usually the risk that can only be accepted are risk that are difficult to handle but considered to have minor impact.

However, accepting a risk does not mean that the risk will be ignored completely. It will still be taken care but with the mitigation strategies. Ignoring the risk without any proper preparation will endanger the company and eventually the business. Even if the risk caused damage to the system or the business, the people will not get panic. Most importantly, the accepted risk must be informed to the whole company to avoid any misunderstanding