1. Objective of Present Work

My idea is to create a new enterprises business application virtual environment to achieve load balancing and failover without investing extra cost or you can say with a non noticeable cost. As I have discussed during introduction that create a new redundant environment for failover is very costly and my approach is based on idea where we can migrate our live running applications from one virtual machine to another in case of the known and unknown outages. Similar to failover in scheduled and non scheduled events, we can achieve proper load balancing by using more than one virtual machines using the high available hardware and software clustered environment. Failover of a virtual machine from one to another would depend on some heartbeats mechanism where the primary and secondary machines will be in synch by sending the heartbeats to each other in a regular interval of 2 or 3 seconds to let other machine know that I am available at the moment and in case primary machine not able to communicate with secondary or secondary found no heart beats in a particular interval then on the based on defined algorithm application will migrate from primary to secondary machine without any outage. In the same way load balancing can be achieved by using different virtual machines and by creating redundant application environments. So without any more extra cost we can achieve the failover and application redundant environment for proper failover and load balancing.

Will discuss in details on this under Methodology section.

So business continuity is the main goal of implementing an Disaster recovery and high availability solutions. High availability is really required if your business really needed the availability over 90%, if availability over 90% is not really required then high availability is not really required but since today every organization information technology infrastructure directly related with the revenue so almost each and every organization needed the availability over 90%. So each and every organization has their own criticality and severity depending on their business modules and considering the same then a high availability can be designed and implemented.

It’s not really very easy to calculate the exact cost with respect to the outages and a user connectivity to the system is not really meant with the system is down. Cost of the outages depend on lots of facts like for how long the system was down, was the complete functionality was down, was a particular module was down, was there any performance issues and system was responding in the regular interval, or was there any issues from client or vendor network side etc. A user connectivity with the system always doesn’t mean that system is not available since there could be your systems are running properly but some performance issues from client or vendor network side.

1) Main objectives of my research work

1. Business application continuity without or with minimum interruption

The primary objective of any highly available business continuity environment is the 24*7 continuity of their business applications without any interruption. Today almost all of the organizations are spending a lot to achieve and build a highly available environment for the
continuity of their business applications because it’s directly related with the cost and revenue of the organization.

2. **Easy and fast monitoring and fast failures detection**

   There should be some perfect monitoring in place to detect the failure of the primary machine where your applications are running like I have been working to develop a solution called heartbeats to communicate between primary and secondary failover machine to know each that primary machine is running or not. So in case of communication got lost due to the failure of the primary machine, all of the running applications from primary machine should migrate or failover to the secondary machine without or with very minimum time frame.

3. **Auto failover within minimum timeframe**

   In case of failure detection, the failover from the primary machine to the secondary machine should be in minimum timeframe to make the proper continuity of the business applications. Today the applications took almost 15 to 20 minutes during failover migration and the business applications doesn’t respond in that time frame but yes, if you have implemented a proper cluster highly available solution then you will have zero outage in such kind of failures because other applications in the cluster will respond in case of the failover of a clustered machine.

4. **Rapid recoverability in case of failure**

   Application continuity or recovery should be in minimum timeframe. Application continuity means a user can able to login to the system and should able to perform the regular operations without any issues.

5. **No session data loss with session replication**

   There should not be any loss of connected data due to the failure of the virtual machine or application servers. Loss of connected data means, suppose a user accessed an application and doing something on that and mid of a transaction and from the backend the connected machine or application server got creashed in that case the user data should survive and replicate to the another machine or server in the cluster.

6. **Create and implement a best planned operational environment**

   During the initial setup of a business environment create a properly highly available architecture with complete failover and load balancing capabilities using the hardware and software highly available cluster services.
Overall you can say the primary objectives of highly available solutions are to provide reliability, maintainability, recoverability, failure or error detection and continuous running business environment.