Proposing an Idea for Multi-Way Data Replication Using a Swarm Intelligence in Current Data Centre Practices

[1]Mandeep Dhalor, [2]Ekta Arora [3] Mudit Khaneja
[1]InformationTechnology, [2]Computer Science , [3]Computer Science
[1]Self Employed IT Engineer., Chandigarh
[2]Poornima Institute Of Engineering and Technology, Jaipur
[3]Drishti Soft Solutions, Gurgoan

***Abstract*— Replication mechanisms are used by various data centres which utilizes database management systems or storage systems intelligence to replicate data to remote or nearby sites. Here taking "Multi Way Replication" a step ahead by proposing a replication mechanism for regenerating a replication site by using "swarm fault tolerance intelligence". In "Multi Way Replication" once a replication site, nearby site or the primary site goes down, we cannot replication the effected site till we are connected to a similar replication site. With "swarm fault tolerance intelligence" we can regenerate the site and hence regain the replication structure to the original.**

***Keywords*—** *one way replication, two way replication, state wide data centre, primary replication, multiway replication*

I. Introduction

Today the data is becoming critical day by day. The venerability of data is becoming higher for which are spending heavily of maintaining multiple copies of same data. In a data centre, the application writes a data through database management software such as MySQL on to storage, but it is not affordable to lose the data and hence we maintain another copy of the same data by replicating it to another nearby site. Nearby replication site can be further be replicated to another replication site which can be nearby or remote replication site. More the number of replications we do, more the number of copies of our data are with us. But keeping every copy of data also costs same amount of storage required. This increases the cost in integral multiples. But having a mechanism to regenerate data from the other copies of data can

- Retrieve the original structure of the replications automatically.
- Can regain the performance in case of distributed load automatically hence acting as a load balancer.
- Brings down the management of data manually for coping it from other copies of data present in other replication sites in case of data failure.
- Brings down the downtime because of self regeneration of data.
- Can also be used to keep a check on the consistency of the data by comparing the data propagation from primary to replication sites.

II. Accepted Replication Mechanism

2.1 One way replication

One way replication is very simple and basic. Here primary copy of the database is replicated to another replication site that is shown in Figure 1. Whenever there is any state of anarchy the replication site is utilized as a secondary alternative. In this method there is no failback that can be done from the secondary to primary.

Figure 1. One way replication

2.2 Two way replication

Two way replication is one step ahead of one way replication as here the mechanism allows us to failback from secondary to primary as well, shown in Figure 2.

Figure 2. Two way replication

2.3 Multi way replication

For more than one replication site we can do replication from primary to secondary to tertiary site or in some other order but this necessarily need to have more than two replication site shown in Figure 3.

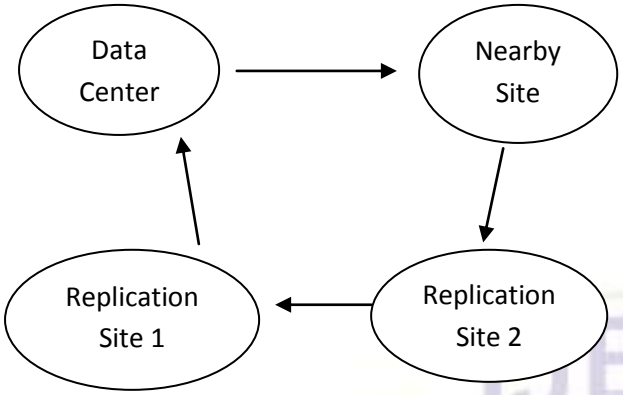


Figure 3. Multi way replication

III. Limitation to Accepted replication Mechanisms.

3.1 There are limitations to the accepted replication mechanisms such as

- Replication cannot be failback automatically to alternative storage because of limited hardware availability. Every time a site or storage goes down we cannot bring another one up until unless resources are not allocated.
- In case of distributed load, there is a performance crunch whenever a replication site goes done which can only be regenerated by building the storage data manually to another site.
- Increased risk and downtime in case of multi site failure.
- No protection against data inconsistency as corrupted data is also replicated along.

3.2 Case Study

In India, every state came out with its own SDC (State wide Data Centre). These Data Centres are for various government and public window operations. For every state there is a SDC which can be referred at www.nic.in. These data centres are geographically spread across India. All these SDCs are either one way or two way data Centres. Problem with the proposed solution:

- In one way or two way replication, there is limited hardware availability. Failback from secondary to primary often requires rebuilding of environment again for primary site.
- Having only two data Centre (Primary and Secondary) limits from having better performance.
- More downtime in case of multi site failure.
- No protection against data inconsistency.

IV. Solution

The solution arrived from one of the natural processes come from SWARM behaviour which can be seen in bees.

- Firstly, on completion of one colony in a bee hive, the young queen and king migrate to some other potential location and start breeding to raise a new colony.
  This behaviour of SWARM can be utilized for guiding replication to happen over multiple similar sites as in case of SDCs which are going to be more than 25 in India.
- Secondly, you will find them to continuously replacing the weak or dead bees for continuous work.
  This behaviour can be utilized to move smaller chunks of permanent errors in database to other storage location.
- Thirdly, on destruction of bees' colony they migrate to another place which can provide them shelter and safely.

In case of emergency or anarchy, if one replication site goes down the process can always start its replication by using other available sites.

Let us assume a scenario when a primary site, secondary sites are tertiary sites are available. The primary site replicates the data to secondary and then the data gets copied from secondary to tertiary site.

Using the first solution of the SWARM behaviour we can replicate the same data from one to many replication sites by fixing the top number of Nodes (alive replication sites) that we want to replicate.

The second and third solution from SWARM behaviour can be copied to make the data available to multiple sites in case of partial or complete data lose.

For Case Study done in Section **3.2,** if all the SDCs are connected together with each other, this will form a network and thus will generate a perfect use case for the proposed theory. If we fix the number of nodes to be four, there will be a primary site, then a nearby site, a replication site and then another replication site. Data will flow from primary to secondary to first replication, then from secondary to second replication site and so on (for node greater than 5). Whenever any replication site is affected, the SWARM replication mechanism can automatically start itself in some other SDC to pull the data from three remaining data sources (Primary, Secondary, replication sites). The regenerating method of swarm behaviour as shown in the following Figures 4,5,6,7 which shows replication of data whenever there is a situation of anarchy.

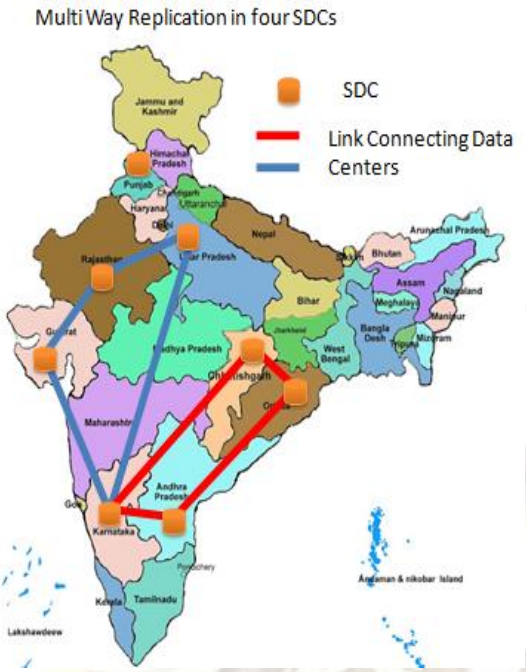


Figure 4

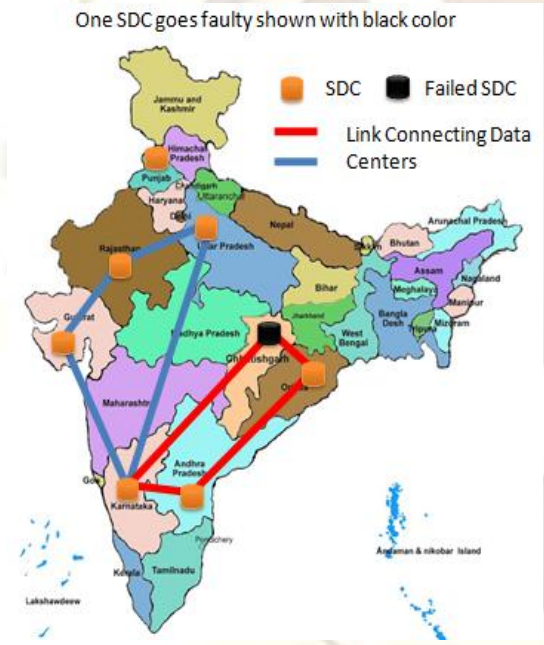


Figure 5

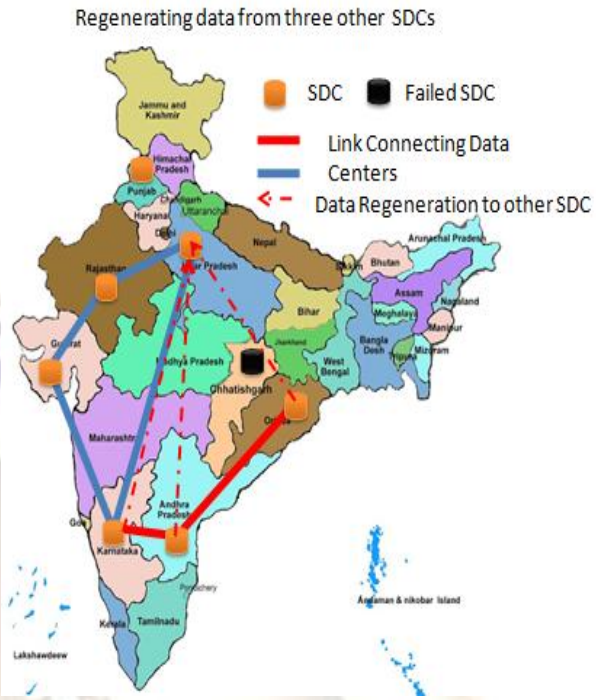
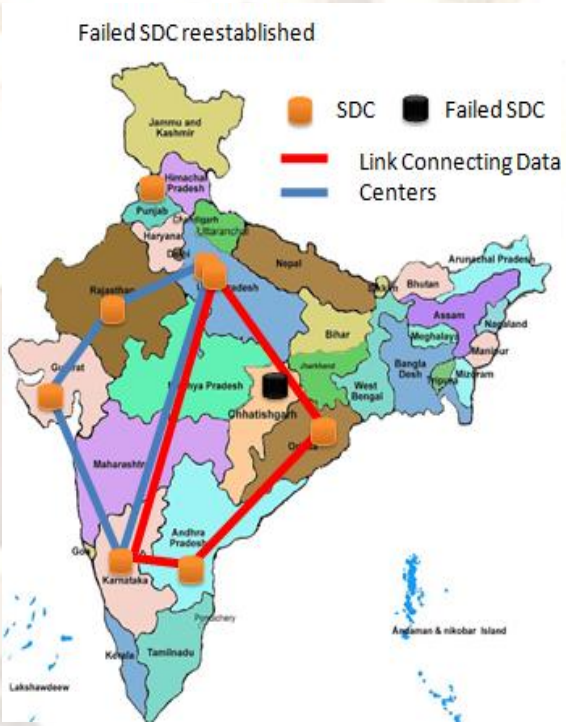


Figure 6



5. Conclusion

With the study of the SWARM natural behaviour, data can be replicated autonomously across replication sites with proposed mechanism. Based on SWARM intelligence it can be replicated and regenerated as well at different replication sites.

References

[1] Oracle web page: http://docs.oracle.com/cd/E22289_01/html/821-1272/multiple-data-Centre-topology.html
[2] IBM web page: http://publib.boulder.ibm.com/infoCentre/db2lw/v8/index.jsp?topic=/com.ibm.db2.udb.replctrdoc/replctr/about3tierrepl.htm
[3] MySql web page: http://www.mysql.com/news and-events/generate-article.php?id=2008_14
[4] Swarm behaviour from webpage: http://en.wikipedia.org/wiki/Swarm_behaviour
[5] Data pulled from: http://deity.gov.in/content/state-wide-area network-swan and http://www.nic.in/projects
[6] A.F.T. Winfield and J.Nembrini,"fault tolerance in robot swarms" Int. J. Modelling, Identification and Control, Vol. 1, No. 1, 2006.
[7] B.D. Shirodkar , S.S.Manvi and A.J.Umbarkar " Multicast Routing using Swarm Technologies" International Journal of Recent Trends in Engineering, Vol 1, No. 1, May 2009