Annexure - 4

Best Practices for DR and BCP Strategies:

1. The National e-Governance Plan has identified multiple mission mode projects along with various e-governance initiatives at state level to provide increased number of services electronically. The State Data Centres would provide common IT infrastructure to host e-Governance applications. Considering the criticality of data/information and impact on the services and revenue in case of any disaster, it is required to have a Disaster Recovery and Business Continuity Plan for State Data Centres.

The primary objective of disaster recovery planning is to protect the organization in the event when some critical or parts of its operations and/or computer services get disrupted or become non-operational. To what an extent and duration acceptable within which Data Centre operations can be revived either fully or at least for critical or minimal required functions, the scale and complexity of the Disaster Recover option would have bearing on investment.

2. A broad analysis of various Disaster Recovery Strategies for State Data Centres are described here under:

Following are the broad factors which would be considered while deciding upon a Disaster Recovery and Business Continuity Strategy:

- Risk Mitigation & DR approach
- Levels of availability or Business Continuity
- Contingency Planning - Risk Analysis
- Cost of downtime
- Application Criticality & SLO/SLA documents
- DR Budget Estimation
- Determining the acceptable cost

- Risk Mitigation & DR approach

The Recovery Point Objective (RPO) and Recovery Time Objective (RTO) are defined by the Disaster Recovery Institute International and are also known as recovery and freshness windows.

RPO defines how fresh the data is after recovery of systems on DR site (second site) after a disaster strikes on production site (first/primary site). Based on RPO definition for a particular application/business/department/enterprise/datacenter, acceptable limits of data loss can be arrived at & thus one could choose necessary technologies for data replication to DR site, mode of replication, quantum of investments requirement to support required RPO definition etc. Better RPO would require high level of technologies to reduce the data loss.

RTO defines how fast, after the disaster strikes on production site (first/primary site), one could recover and have a particular application /business /department /enterprise /datacenter running. Based on RTO requirement/expectation of users, suitable technologies are deployed to ensure recovery within the defined period. RTO definition would influence the speed & time in which DR Datacenter should be able to resume a particular application/service to the user. Better RTO will require higher technologies and efficient recovery processes to have speedy recovery.

- Levels of availability or Business Continuity

Different levels of availability or business continuity which are possible for each application or service are as below:

- Basic availability
- High Availability
- Disaster Recovery (DR)
- Disaster Tolerance (DT)/ Business Continuity (BC)

Depending upon what RPO & RTO a particular application or service needs, one could chose the appropriate level of Availability or Business Continuity.

BASIC AVAILABILITY is what is built with in a system (could be a hardware or software unit) and is known to provide first level of availability. It protects from component or subsystem level failures in a single system. This assures basic uptime to a system and costs the least among the four options under consideration.

HIGH AVAILABILITY is the next higher level of availability where the application or service continues to function even if there is a system breakdown. This availability solution aims to eliminate as many Single Points of Failure (SPoF) within a system/deployment architecture. High availability is normally under the same roof.

DISASTER RECOVERY (DR) allows an application/service to be recovered on an alternate site away from primary production site in case a higher level of failure or disaster strikes, resulting in production or primary site becoming completely unavailable. Recovery process takes time before the secondary or backup site is made operational. Delay depends upon type of technologies deployed or recovery approach taken. DR approach may have many possible options to achieve disaster recovery. A simple example is to have a remote storage of backup cartridges in a remote place or DR site. A sophisticated (and expensive) approach is to have network connectivity between the two sites for online data replication and where DR site also has a minimum infrastructure to ensure data replication operation. (These are two extreme ends of DR approach in terms of RTO/RPO definitions.)

In the above first approach, backup cartridges are shipped to DR site and just kept safe. In case of a disaster striking on primary site, necessary system infrastructure (servers, storage, software etc.) are organized and then data from

backup cartridges is restored to recover the site. This is a low cost approach but recovery time is high, sometimes few/many days (poor RTO). Also freshness of data, after recovery, is poor as it can recover only till the time of last good backup available, which may be a day or more old (poor RPO).

In second approach, DR site has some minimum necessary infrastructure deployed and typically incremental Data updates from Primary are shipped over network either in real time or at regular intervals. The replicated data updates are stored in a manner exactly similar to its Primary counterpart. In case of a disaster at Primary, services can get started out of the DR site. This approach is costly but provides quick recovery (good RTO). Here freshness of data, after recovery, is much better (better RPO). This approach may use various data replication technologies in conjunction with RTO & RPO definitions of the organization/department.

DISASTER TOLERANCE (DT) or BUSINESS CONTINUITY (BC) ensure continued operation without requiring any/considerable delays that normally happens in DR. Theoretically, it tolerates the disaster and operations start on the second or backup site. Actually DT site quickly takes over the operation with minimum delay and there is hardly any significant and time taking recovery. It requires use of best in class technologies to have best possible RTO & RPO. In many cases, DT sites deploying capacity nearing to 100% of the production site capacity to ensure no significant performance and capacity degradation after primary site failover. It always has real-time data replication to DT site with minimum or no delay.

- Contingency Planning - Risk Analysis

Contingency planning is an activity that strives to ensure that an event doesn't become a disaster. It doesn't just cover a computer system or the Information management department; it covers a wide range of business and technology issues. The objective is to ensure that a Department's/ Organization's

business processes are recoverable and that valuable information is always available.

As a part of risk analysis and to protect Department's/ Organization's business, it is important to know "what organizations are trying to protect":

- What are their key business processes?
- What are the threats against those business processes or in another word what are they trying to protect these from?
- How much will it cost them if those business processes stop and how much are they willing to spend to keep them from stopping in terms of time, effort and money?

The purpose of a risk analysis is to answer the above questions. This is the foundation of DR planning.

Business risk is a function of the likelihood of a disastrous event occurring and its potential impact. Both need to be assessed. The likelihood of an occurrence is itself a function of factors that include:

- the availability levels of the current infrastructure
- hours of operation
- the chance that downtime will occur during normal hours of operation

The factors affecting likelihood, if viewed in a little greater depth, one could realize that a system that is available 99% of the time will experience nearly 90 hours of downtime during a year, while a system that is available 99.999% of the time is down less than five minutes during the same period. 99% availability may have a significant negative impact if downtime intersects with peak usage periods during a "standard" 8 hour business day or if it operates an e-business that must be available for longer periods.

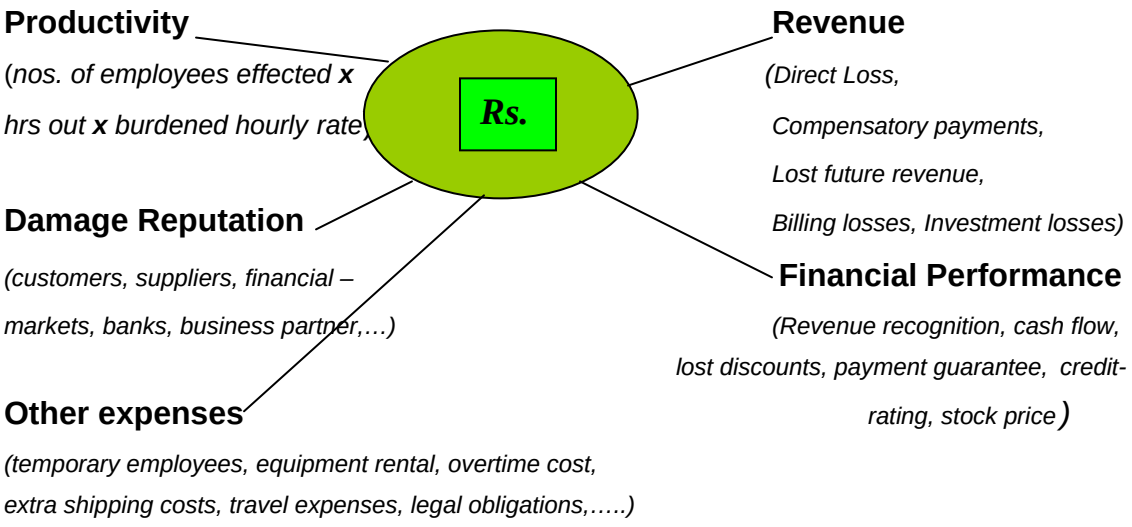
- Cost of downtime

As business and IT managers, one needs to understand and analyze cost of downtime. Overall loss to the organization/department due to downtime can be divided into two categories i.e. visible loss and Invisible loss. Loss in revenue due to operations & services coming to standstill, as a result of disruption/downtime, can be treated as visible or direct loss. Invisible loss is difficult to determine but it is known to have long term effect. Under invisible loss, one may consider loss of productivity, customers, reputation etc. which is normally more than the visible loss.

So while working on cost of downtime, the above factors (Visible & Invisible) may be taken into account.

The cost of infrastructure, facility, staff, training, consulting and other services may be considered while arriving at cost of high availability or business continuity.

The cost of downtime: Implies the followings



(Down time cost per hour, day, two days, …….)

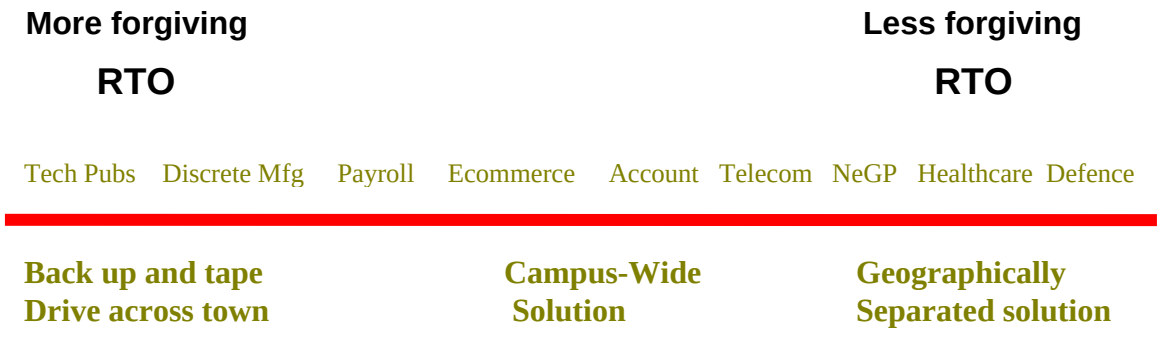
To set the RTO, it is required to know how much downtime costs. To measure the cost of downtime, it is desirable to determine how much availability one

requires. Is a particular application in the Datacenter “nice to have” or is it “business critical” or “mission –critical”? What is the impact on the department or on the Datacenter - of an application failure? Can it be measured in lost productivity? Lost revenue? Lost profits? Damage to reputation? What are user expectations or regulatory requirements? One way to assess the cost of downtime is by “walking through” a sample disaster, such as a fire in a data center, and assessing the potential impact. Once a thorough assessment is done, multiply lost productivity and revenue per hour by the number of hours to recover the current infrastructure to arrive at the cost of downtime.

- Application Criticality & SLO/SLA documents

Application criticality: Any business recovery or business continuity plan must begin with an understanding of the application requirements. How often does the application or suite of applications run? Is some downtime acceptable? How critical is the application? How often should you back up data—continuously, every minute, every hour, once a day?

The business unit that needs an application usually defines application criticality. This definition is based on whether loss of access will be detrimental to the business unit. The RTO is how much time users can afford to be without the application.



SLO/SLA documents: Organizations often use a document called a service level agreement (SLA) to define various service level objectives (SLOs). SLOs include criticality, acceptable downtime, normal hours of access, number of users,

acceptable response time, and other factors. Business needs determine the actual values associated with these SLOs. With the appropriate level of business recovery or business continuity the service level objectives can be ensured and can be achieved within the recovery time objective.

- DR Budget Estimation

Budget required for DR plan may vary upto a great extent depending on variables such as criticality of applications/data, level of security required, risk factors, RTO (Recovery Time Objectives), RPO (Recovery Point Objectives) of the organization concerned. Disaster recovery systems can restore access to the application in minutes, hours, or days. The biggest difference is cost: how much investment is needed for additional data centers, hardware, software, and communications links. Another major aspect of cost is people: how much one wants to invest in training, developing redundant skills, creating procedures, and fostering communication.

From a investment (for DR Datacenter) perspective, bottom line is that more data loss a business/service can withstand, and the longer a user/customer can wait to recover, the less it will cost. However, the longer a business process/service is unavailable, chances are that greater would be the losses (financial/ reputation/ user satisfaction etc). There is a point at which the potential loss equals the cost of recovery. This is one way to determine how much to spend on the plan and how long the service/process can be allowed to be unavailable.

Based on the basic criterion for business continuity and DR as discussed above, the estimation for DR planning may vary from as low as 35 % to as high as 80 to 90% of Primary Data Centre cost.

- Determining the acceptable cost

While these best practices discussed above would be meaningful & relevant on a broad scale to any organization/department who’s planning to setup a DR

Datacenter, a more specific plan & cost estimate can only be prepared by taking a more closer look into the key attributes of the Primary Datacenter.

One has to analyze each and every application/service being hosted out of the Primary Datacenter, determine its criticality, RPO, RTO & then accordingly map - in the application in the Datacenter's Service Level Objective (SLO). Thereafter, based on the criticality rating of the application, one can proceed further towards arriving at what DR investment would be justifiable for the particular application/service. All such application/service specific DR investment estimates would cumulatively lead to an acceptable DR cost.

3. Various options that States may exercise to build their DR Data Centre are as under:

i. Regional Disaster Recovery Centres (RDRC)
ii. States create their own DR plan and facilitate it independently.
iii. State Data Centres acting as DR site for like minded States
iv. Outsourced Model for Disaster Recovery and Business Continuity