

Continuous Data Protection

By David van Geilswyk

When we think of a high availability solution that provides a business with disaster recovery protection we are typically thinking of a configuration where the source and target systems are geography separated running real time replication. We measure the effectiveness of this solution in terms of recovery time objectives (RTO) and recovery point objectives (RPO). In business terms these translate to:

- **RTO** = How long can you afford to be down?
- **RPO** = How much data can you afford to lose?

With this type of high availability solution in place we still need to do our point in time backups (nightly, weekly and monthly) in order to protect ourselves from data loss. Data loss can come in the form of software corruption, security breach, viruses and accidental deletion. A high availability solution does not protect us from these types of problems as it would replicate the data loss in real time to the target system. The data loss exposure that we face even though we have a high availability solution in place is that we would have to go back to our last backup of the data for recovery. Depending on the extent of the data loss the business's recovery time objectives and recovery point objectives may not be met. Continuous data protection (CDP) was designed to address this exposure.

Continuous Data Protection

CDP solutions automatically save data modifications independent of the primary data which allows the administrator to recover from any point in time. The implementation of CDP works along side point in time backups and replication to improve recovery time objectives and recovery point objectives for your business. Since CDP saves data modifications, recovery point objective from data loss is driven towards zero. The recovery time objective is improved since there is a full copy of the data prior to corruption, though depending on the extent of the incident, size of the recovery window (more on this later) and any analysis required, some time

may be required before normal operations can resume.

To run CDP, disk space is required to support the second copy of the data with additional space required for the window. A small amount of memory will be required for the process and the processing requirement for CDP is minimal. CDP is meant to complement a high availability environment and should be configured along side it. CDP provides the business with a window of opportunity to detect and correct any corruption.



David van Geilswyk

How Continuous Data Protection Works

For an example of how CDP works we will examine Lakeview Technology's iSeries solution contained in both the MIMIX HA1 and MIMIX HA Lite products. Lakeview Technology a leader in the high availability space. CDP works by setting a recovery window which is the amount of time required to detect a data problem. Another way of looking at the recovery window is: "How far back do you need to go?" The software allows for a recovery window of from 1 minute to over 60 days. What the recovery window does is stop the apply process to the CDP target area. That ►►

MIMIX

Typical HA Configuration

Site: Toronto
Role: Source



Site: Montreal
Role: Target



means you will need to start the apply process and wait for it to reach the point just prior to the corruption before you would be able to use the CDP data. Lakeview has another term called the availability recovery window. The availability recovery windows is essentially the recovery window plus any lag in the apply process.

In the recovery window we can set recovery points which can be immediate, date/time or by journal

Turn Your Sow's Ear into a Silk Purse.



Is some overpaid consultant telling you to throw out your System i (AS/400) applications and replace them with a "prettier package"?

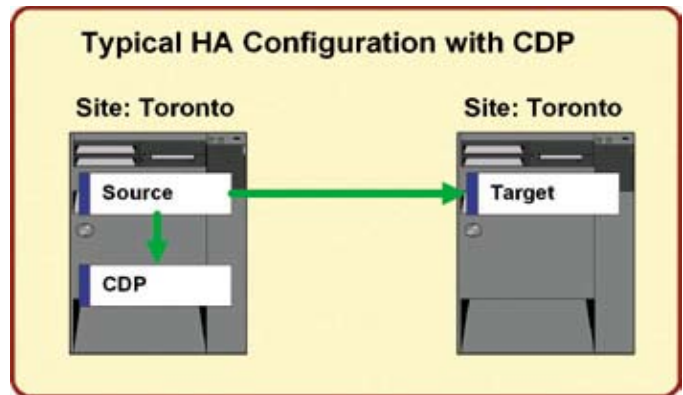
Tell them:

- A. You have considered the costs of overpaid, self-serving consultants.
- B. You have considered the soft costs such as staff re-training, (the costs the overpaid, self-serving consultants ignore) to do so.
- C. You have considered the hard costs and moving to Wintel is full of them.
- D. You have considered the costs of modifications to make their "pretty stuff" actually work in your business environment.
- E. All of the above.

....and tell them you have decided to pilot **Lansa Ramp** from **Mid-Range** because you will get twice the results for less than half the price.

Note: When your self-serving, overpaid consultant hears this, he or she may switch to Plan B and borrow your watch to tell you the time. Don't let them.

For more information please e-mail us at lansa@midrange.ca.



entry. The recovery point is the point at which you suspend processing for a data group for the purpose of recovering data before corruption is replicated. Once a Recovery Point is issued no further processing will occur until it has been cleared or restarted. After setting the recovery point, the apply process will continue until that point has been reached. Upon reaching it there are a number of options available. What to do at this point will depend on the extent of data loss and its impact on the business. Here are some of the options available to us at this point:

- Take a copy of the CDP data and move it into production, then restart the CDP apply process and let the resolution flow through.
- Promote the CDP data to production.
- Skip over the corruption in the apply process, let the apply process complete, then promote the CDP data to production.

Closing Thoughts

We see the business case for CDP being particularly strong in environments that have high rates of data change where a loss of 24 hours worth of data would be crippling. It would also be useful in environments with fairly large amounts of storage as this type of environment could take some time to restore. The solution provides a definite advantage when downtime would have a significant impact on the company.

CDP provides us with a solution to a long standing high availability issue by protecting us from data corruption that flows from the source system to the target system. CDP can be run outside of an availability solution, but the real value comes in running it in conjunction with a high availability solution. Since more businesses are operating 24x7x365 and business today has less tolerance for day loss then ever before, we see CDP being adopted by a large number of high availability users in a relatively short period or time. Thank you for taking the time to read this article and I hope it has served as a good introduction to CDP. **M-R**

David van Geilswyk has over 15 years of IT experience on the System i5 iSeries/400 platform and specializes in Back-Up and Recovery, BRMS, High Availability, Systems Security, and Logical Partitioning. David can be reached at: dvg@midrange.ca.