

Technology Overview:

Replication for the Masses

By Steve Kenniston

IT professionals have increasingly become certain that in order to ensure company longevity, data must be stored in multiple locations; this is not only “nice to have”, it is a necessity. In addition to secure, reliable backups, IT professionals are looking at replication as the next killer application to make sure that their company’s data availability metrics are met on a daily basis.

The ability to understand both what needs to be replicated, and why the data needs to be replicated, are the two most important questions to answer when embarking on a replication project. Once these questions are answered, selecting a replication technology becomes easier.

There are a number of reasons to replicate data; Business Continuity, Disaster Recovery and real time data access for reporting are just a few. Narrowing down the specific reason for replication will dictate not only the type of replication that is needed to solve the business problem, but will also help when making infrastructure decisions such as what bandwidth is required to replicate the data.

Once the “why” question has been answered, the next important decision to make is what data needs to be replicated. Knowing why data replication is important will help to determine which data in the environment needs to end up at a remote location. Some simple storage resource management (SRM) tools can be used to help understand where the data that needs to be replicated currently lives. Chances are that while building a replication strategy, some infrastructure changes may need to take place. Identifying where the data sits can help to determine where some infrastructure changes may need to take place and will ensure proper budgetary planning as a replication strategy takes shape.

Next, in order to round out the replication strategy, selecting the type of replication will aid in determining the band-

width requirements necessary for replication. There are two fundamental types of replication:

- Synchronous
- Asynchronous

There are a few other ways to replicate data; vendors spin these other ways for replicating data differently to express the value of their philosophy on replication as well as their advantages. In the asynchronous mode, technology such as “copy on write” is used where by the full data set is not replicated until there is a request at the remote location for the data. Data that is not required at the remote location does not eat up bandwidth. Another method used to replicate data in asynchronous mode is by writing meta data to log files and not disk when passing along a confirmation. However, at the end of the day, these are still the two basic ways to replicate data; synchronously and asynchronously.

Synchronous Replication

Synchronous replication requires data be physically written to both arrays, and both arrays must issue a confirmation that the write was successful prior to a second write taking place by an application. The benefit of synchronous replication:

- Very consistent and accurate data on both arrays

The down side to synchronous replication:

- Performance issues
- Costs (due to performance issues)

Asynchronous Replication

Asynchronous replication does not require a write confirmation from the secondary array before a new data block can be written to the primary array or sent off to the secondary array by



Storage Knowledge - Business Results



the application. Typically, the next write is sent when the first write has made it to cache on the secondary array. Benefits of asynchronous replication are:

- Better performance
- Cost effective

The down side to asynchronous replication:

- Greater potential for data inconsistency

The type of replication desired is dependant upon the business problem that is being solved. Replication for disaster recovery with the sole purpose of having a secondary copy available when or if the primary site goes away can take advantage of asynchronous replication. Business applications that are run 24x7 with reporting for “up to the minute” decisions would need synchronous replication.

There are two primary methods of replication and a third emerging method. Array based and host based replication have been the primary replication methods of choice. However, innovations in appliance-based replication that are migrating into a fabric-based approach and moving even further into the switching layer could provide an additional non-traditional approach to replication with a lot of upside.

Array Based Replication

One of the predominate methods of replication today is array based replication. Array based replication consists of a pipe between two disk subsystems. This is typically a dedicated pipe and both subsystems must be alike. Usually firmware is installed within an array and it controls the I/Os from one array to another without taxing the server(s) they are attached to. Array based replication is usually very reliable due to a dedicated pipe, which conversely can be very costly. In addition, deployment can be expensive with consulting fees.

Host Based Replication

Host based replication consists of software installed on a host that controls I/Os from one server and array to a secondary server and array. Host based replication tends to be more affordable than array based replication and allows replication between heterogeneous disk subsystems. In addition, host based replication does not require a dedicated pipe between disk arrays. Host based replication tends to be a little more difficult to implement because, unlike array based replication, there are more parameters to take into account. Newer host based replication technologies are making it easier to implement this software based solution which is particularly helpful for first time deployments today.

Fabric Based Replication

Making its way to the forefront are replication appliances and intelligent storage switches. The intelligent switches have technologies in the switch that can do volume management, virtual tape library emulation, and replication. The benefits here are a combination of both host and array based replication. Moving replication into the fabric allows users to have dedicated lines from one array to the other; it also allows for array heterogeneity and good performance as I/Os do not need to go all the way back to the host. By having the software in the fabric, implementation becomes easier as well.

The last component vital in replication process is the pipe between the arrays in which replication is going to take place. Bandwidth can be the most expensive component to the entire replication solution. A synchronous replication solution typically requires a dedicated pipe between arrays so that performance is reasonable for the application and ensures consistency.

At this point, the business reasons for replication and the data required for replication are determined. The final decisions on the ways in which to replicate come down to what one can afford. Determining the cost of down time can help determine if synchronous or asynchronous replication is the right choice. In most situations, asynchronous replication meets all the business requirements for replication. Host based replication lends itself well, for the cost, to asynchronous replication. Additionally, an interesting side effect that is very common with replication implementations is a company's ability to move up the business continuance ladder. Traditional backups lend themselves to a decent disaster recovery strategy, however, combined with replication, a company's ability to close the backup window all together as well as provide faster access to data in the event of a disaster help move companies from traditional disaster recover and closer to business continuance.

Finally, make sure to set up proper best practices for replication. Fundamental checks on the remote data are important to ensure that the data is accurate. Testing the amount of time it takes to ensure users can run off the remote set of data is important when establishing if the replication type selected meets the business need. The ability to migrate users back to the primary site is an additional test required to ensure that the setup is correct. Performing these tests on a regular basis will ensure that when an issue does arise, everyone is trained on the proper processes and procedures to ensure success and the objective of keeping the business running is met.

All trademark names are property of their respective companies.

Information contained in this publication has been obtained by resources the Enterprise Storage Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time.

This publication is copyrighted by the Enterprise Storage Group, Inc. and is intended only for use by subscribers. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Storage Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Please contact us at (508) 482-0188.