

BUSINESS INTELLIGENCE

with Jackie

Combing High Availability with Data Warehousing



Jackie Jansen

“High Availability and Data Warehousing, a match made in heaven”. If that is a bit of a stretch how about: “Two technologies that were meant for each other”. All teasing aside, I will tell you that from a business point of view these two solution areas really do fit extremely well together...



If you have a mission critical application that drives the requirement for a high availability system then

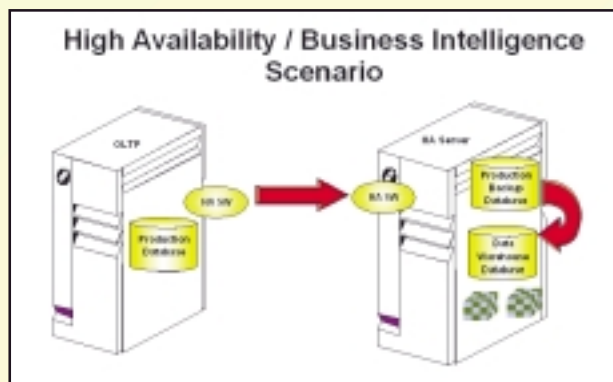
justifying a data warehouse just became a whole lot simpler. If you would like the security of a high availability box, and after the recent worldwide events who wouldn't, then possibly data warehousing may give you the extra ROI you require to cost justify an additional system.

Let's start by looking at this from a high availability point of view. An obvious requirement for high availability is that you replicate your mission critical data from your production iSeries or AS/400s onto a backup server. This backup server has to be available at a moment's notice to take over should your primary production server experience a failure. What you don't want, though, is to have your backup server sitting idle. A data warehouse or Business Intelligence solution is often not considered to be a mission critical application when it is first implemented and is an excellent choice to utilize the additional CPU cycles.

Does having your data replicated in production format on a second server

mean that you have a data warehouse? The answer to that is a resounding NO! What you do have is system where you can run your operational queries and reports without impacting your production system. You also have a system that contains the source data that you need for your data warehouse and presumably quite a few available CPU cycles.

Now from a hardware perspective is this a “free” data warehouse. The answer is typically not. Although you will usually have enough CPU cycles to run your




data warehouse you will still need to add additional disk space. As mentioned earlier, what you now have is the source data for your warehouse. One of the characteristics that differentiates a data warehouse from your operational system is the database design required for large, potentially unstructured queries to occur efficiently. You will still need to

transform and load your data but you will no longer need to impact your product system extracting the transaction data. This can all be done within the one server, directly from the replicated version of the data.

Typically production systems have a nightly window that much of their batch processing has to occur within. This window is often very tight. A large advantage to using the high availability server as both the source and target system for your data warehouse is the elimination of this tight batch window.

Although you will still have a schedule that you are trying to keep to, you are no longer competing with all the end of day jobs running on the production system. As an aside, you should realize that you can also use the high availability server to do your nightly backups and your end of day reporting.

Two technologies meant for each – High Availability and Data Warehousing. 

Jackie Jansen is a Certified Consulting IT Specialist. She currently works in the IBM Americas Business Intelligence Solutions Centre. Jackie is a frequent speaker at AS/400 Technical Conferences and User Group meetings. Contact her at jjansen@ca.ibm.com.