Computer Consulting and Support Group Notes—Special Meeting on October 12, 2009

This is a special joint meeting of the IT coordinators group (which specializes in the administrative side) and CCSG (which is more related to the academic side) to discuss the recent outage on Exchange server EXCHS017. People in both camps were affected, so we brought everybody together today.

**Presenters:**
- CIO Jim Davis (opening comments)
- Mike Lohrbach, leader of the Windows Applications and Servers (WAS) group, on technical aspects and future plans

**Jim Davis**
We thought it was important to get everybody together; the crisis is over but fresh in everyone’s mind. We’ll try to answer two important questions: what happened, and how will we prevent this in the future?

We know that you are out there helping us deal with this. We thank you for helping us communicate appropriate information to those impacted.

Our goals for today are to talk about where we’re at and where we’re going, and to hear about what you were seeing, what worked, what didn't, and what we need to do next time.

**Mike Lohrbach**
As head of the Windows Applications and Support group, Mike was our point person on the response team for most of the outage.

We do appreciate all offers to help us out and the feedback as to what users were seeing.

We became aware of an issue on Wednesday, September 29. We found a failed disk in one of the storage groups for the Exchange server EXCHS017. That caused an issue with the information store service, and everyone lost connection with that service (resulting in inability to access mail, calendars, etc.). We replaced the hard drive and started a rebuild, which should have fixed things. However, we had the service fail two more times.

From 5 p.m. that Wednesday until about 11 a.m. on Thursday, that server was stable. Then the information store service failed again. At that point we started looking at different options. One group was looking at fixes for the existing data store, while another looked at migrating users to another server. We finally decided to migrate users to another server that evening. However, the information store service wasn't stable enough to migrate users.

We were able to isolate the problem to one info store. At 11 a.m. on Friday morning, we removed that store from the server, and mail service became available to other users not
on that store. At that point we contacted Microsoft support and, at their recommendation, we started running the repair utilities they supplied on the problem info store. That finished early Saturday morning. We brought the repaired store back on line, then attempted to back up the server. The backup process checks for corruption; it returned errors not on one, but on all three databases. We then started repairing all three databases, hoping to get it done over the weekend. That process finished at about 9 a.m. on Monday morning. After that repair, the databases stayed mounted but the local delivery queue wasn't working.

At that point we decided to attempt a restore from backup. The biggest obstacle to all of these processes is that these are big files; the three databases amounted to about 700 GB of data altogether. Moving that much data or scanning large data stores takes a lot of time.

On Monday evening at 2 a.m., we started the restore. The first data store restored at 8 a.m. on Tuesday. We attempted to mount the store, but it didn't mount correctly. We then started working on the other two data stores, but the process wasn't working correctly.

By this time storage group 1 is online, but group 2 wasn't. We then determined that we needed to get higher level support from Microsoft. They recommended something that we didn't know about: creating a "dialtone database", which creates new blank mailboxes from the Active Directory information for users. That happened somewhere around 12:30-12:45 p.m. on Tuesday; users could then send and receive mail but not see old mail. Then we had to run several utilities against the old message stores, and that ran from Tuesday to early Thursday (when the first data store, for FPM, was finished). Marc Witte helped us a lot here by getting us information on what his users were seeing.

Friday morning we brought message store 6 on line and started the merge process. There were still some issues with calendars and Blackberries (many of which had to be completely reset). Windows Mobile and iPhone devices seemed to come online correctly.

Saturday morning we did the merge processes on the final database, bringing the last users online.

That's the overview of steps. We learned a lot, it was not fun for anybody, and it was not a good thing. We will put forth every effort we can to make sure it doesn't happen again.

Our focus was to restore service to users as fast as possible. Things we could have done differently:

1. Watch database sizes. There isn't much in place in our current Exchange system to provide for high availability. In particular, Exchange 2003 must restore on top of the existing database; Exchange 2007 has other options.
2. Use as many storage groups as you can. 2003 supports 20 groups per server. Smaller user sets give smaller, easier-to-manage data stores, with fewer people affected when they go bad.

3. We now have new performance monitoring tools we can share with others.

4. Knowing what we know now, we should have gone to the dialtone database sooner.

5. Better restore procedures; EMC is working with us this.

We worked with someone at Microsoft that specializes in disaster recovery. We learned what measures other companies have used to minimize problems, but not all of them would work here. We have diverse clients and varying mailbox sizes (up to 14 GB of mail).

6. Use Exchange cached mode; users could see old data even when offline. Many places use this as recovery processes; they simply leave the temporary mailboxes online and let the users merge data back from the cached folders.

We are focusing on building a highly available Exchange 2007 environment.

We are using clustering to build in redundancy. There are multiple client access servers (which provide Outlook Web Access, IMAP, and Outlook Everywhere) instead of the single exchange.iastate.edu server. This will require migrating the certificate to the new server. The CASs are load-balanced and redundant.

Hub transport servers route mail, they are also load-balanced. The decision was made to go with Microsoft CCR; in that environment we will have two independent copies of the mailbox stores. If you get corruption on one side, you don't necessarily have corruption on the other. There will be one server in ASB and a server in Durham, and all of these will be virtualized.

The disk for these servers will be in the SAN. We are limited to one mailbox store per storage group. (In Exchange 2010, storage groups go away.)

With changes to the way backups are being done in the 2007 environment as well as technological improvements to restore options in Exchange 2007, we would have had more options for recovery of mailboxes.

Backups will be LAN-less; the backup procedure will take a snapshot of the ASB nodes and copy data directly to disk. We will not have to go across the network to get those files back. The snapshot is a sub-second transaction so the impact on the server should be minimal. They'll also be file-level data saves; we'll be able to get data out without needing the Exchange server running.

In Exchange 2007, you don't have to restore to the same storage group. You can actually restore a mailbox to another server, making it easier to move users to functioning servers.
We'll also keep database sizes smaller, with tools to monitor size and server performance more than we have in the past. The limit on storage groups is now 50 per server instead of 20.

Tomorrow we'll be working with Microsoft on a design review of our 2007 environment; we hope to begin moving users into that very soon.

Questions and Answers

Q: Do you know what the cause of the initial failure was?
A: We're not 100 percent certain but it looks like when the disk went bad, the database became corrupted.

Q: Was the problem hardware or software?
A: It appears to have been initiated by hardware, then the corruption had to be repaired.

Q: Is your focus on moving to 2007 rather than fixing 2003?
A: We've discussed that. Unless you can completely remove everyone from the database, removing users doesn't automatically shrink the database. You have to take it offline and defrag it, and that takes as much offline time as scanning it. Our goal now is to focus on migrating to 2007 and validating our 2003 backup and restore processes.

Q: What is your time frame on migrating to 2007?
A: We have only one technical question to get answered: what effect will migrating the certificate have on all the clients connecting to it? Until that actually happens, we won't be sure, even though we're doing testing in a lab environment. Timeline: We were hoping to do internal users that last week. Now...we hope to do so in the next week or two.

Q: Will we keep on virtualizing servers?
A: Unless we see issues that contradict that, yes.

Q: What happens when the next disk fails?
A: That shouldn't happen, they're 14-disk RAID arrays. Microsoft thinks that the problem had to do with the database sizes. Issues are related to performance and the amount of time required to restore and repair. The maximum size of a database varies; it's based on service-level agreements and maximum permissible downtime.

Q: Why does using smaller data stores help if a disk fails.
A: If you have many small stores on a disk that fails, it will just take as long to get everyone online, but some users will come online faster.

Q: If you have two copies of mailstore, how does synchronizing work in 2007?
A: It's not syncing; they’re two parallel servers. The passive server is isolated, updated from the original server logs. It's not mirroring from database to database; the transactions are applied to both databases simultaneously. We had a logical problem with the Exchange database. Had we been using 2007, we would immediately have mounted the backup database and continued on from that point.
Q: What do I tell my users? What are the advantages and what is the expected maximum downtime in future?
A: In theory, the failover from active to passive server is supposed to take minutes. You can even do an automatic failover. Even the worst-case scenario should be less than hours.

Q: What my users were doing was changing forwarding to the POP servers and reading mail via WebMail. If the POP servers aren't available in the future, how do they get their mail? Will ITS provide a set of guidelines for what to do when Exchange is unavailable? In my department, we're recommending that they have another mail account not related to the university for a backup.
A: With what we're doing with clustering and what we've learned about dialtone databases (making it possible to send and receive mail), downtime should be much shorter.

Jim Davis: We felt that we couldn't recommend that people get Gmail accounts for FERPA and other security reasons, but we're reopening those discussions with Google and other interested parties.

Q: We appreciated the updates so we were able to inform our clients. Is there anything in the new Exchange that would pre-emptively detect corruption before it would become a problem?
A: There are some things that can be done. Some of these things are done by the backup process. We'll be watching that more closely. We're going to put a lot more monitoring into the process. We've gotten some good info from Microsoft and are looking at System Center Management. We are going to be as proactive as we possibly can.

Q: Every one in my department was patient but you could tell they were frustrated.
A: I'd like to thank Jim Twetten in Communications for getting those updates out; it was great that he would write and post that information for us.

Q: Are there plans to go to Exchange 2010?
A: We're going to start looking at that as soon as we have 2007 up and running. It's supposed to reduce I/O load on the servers, and the 2010 OWA client is supposed to be browser-independent. The 2007 version is much better, but still works best on Internet Explorer.

Q: What could have helped us is more warning before creating the temporary accounts. We had a lot of people who suddenly didn't have any old mail without sufficient warning.
A: That was a decision we made because of how long we'd been out. We weren't quite sure what issues there were but we wanted to restore email functionality.

Q: Why didn't calendars work under the temporary account?
A: The calendar permissions are tied to the account, and those couldn't be re-created easily. What users got in a temporary account was like a brand new mailbox. You could have gone back and set up permissions again, but all that was fixed when they
were merged. (If we'd kept the dialtone database as the default, they would have had to recreate the permissions.)

Q: Do you know why the backups failed?
A: No. The logs have been sent to EMC and we're going over it as well to try to determine that. We are going to focus everything we can on not having to do this again.

Comment: The big issue was that we have 2/3 of our people who were working and 1/3 that didn't. The working users were sending mail and creating appointments for those people and didn't know it wasn't working. The information needs to go out to everyone, not just the affected users.

Q: The problem log is great, but it's hidden. Many people couldn't find it and the RSS feed wasn't being updated.
A: Updates aren't being sent to the RSS feed, we're fixing that this week.

Q: There needs to be more publicity of the Twitter and Facebook paths as well.
A: We're also working on that. University Relations didn't want to put info on the home page because that's an external page, not internal, and this was internal information.

Q: I was disappointed that there wasn't an easy way to find out what Exchange store people were on. The Solution Center could look it up one at a time, but it would have been nice if I could have checked all my users.
A: You could use the Exchange Admins tools; it adds tabs to AD Users and Computers. You could use also use BlandADE; it's available through the Windows Admin site. You can also use an LDAP query, but that's pretty esoteric and you have to know what to ask for.