mogulskier@hotmail.com

From: mogulskier@hotmail.com
To: Andy Howell
Cc: mogulskier
Subject: Tiered Storage Discussion and Phases of CAMS operations

Your speed is slow, but it is working. You are not dropping frames, etc.

I will explain another more evolved approach that I've been thinking about (based on lessons learned). The purpose is to make the system more robust/durable but to also allow for a smaller SSD style drive to perform the normal capture/working stuff (Tier-0 or 1), then migrate it to a Tier-2 or Tier-3 level storage for safekeeping. In other words, establish the CAMS Tier-0 storage as the place wear serious work gets done that has the fastest I/O on the CAMS system. Then migrate or replicate it to Tier-2 or Tier-3 for the non-I/O-dependent stuff, like archiving. The Tier-2 level stuff would still be fast enough to do things like confirmation, upload to SETI, and manual analysis of the sessions.

This is all based on tiered storage data lifecycle management and will function to achieve a balance between performance, cost, durability, and availability.

- Availability is achieved in the Tier-0 through Tier-2 storage classes.
- Cost goes down as the tier level increases.
- Tier-3 is in the form of archived (zipped) SubmissionFiles sessions. Tier-3 is less *Available* because the sessions would require unzipping to a higher-tiered storage class 2 or higher before it can be manipulated.
-

CAMS Tiers might be:

Storage Tier	Functions/Actions	Comments
Tier-0 or Tier 1	• Capture, autocal, detection, queuing for upload. • Temporarily move working sessions to SubmissionFiles • Migrate Transmitted to Tier-3	
Tier-2	•	
Tier-3	• The FF*.bin files in the CapturedFiles dir of these sessions are deleted. This reduces the storage requirements of an 8 camera system of around 55 GB per night to an average of 645 MB per night, dramatically reducing the storage by 85 times. • Each SubmissionFiles session is zipped into a single file, further reducing the storage requirements by half. • The total reduction in storage requirements by the Tier-3 logic is over 170 times.	

Capture, autocal, detection, queuing for upload
Uploading,

Here are my thoughts:

- Let's say that you have a fast SSD drive that is only 250GB as drive D:.
- Let's say that you have a large, but slower and less expensive HD drive, DAS, NAS, etc. as drive E:

- We capture to D:\CAMS\CapturedFiles
- Post-capture processing will perform autocal, detect, etc on the D:\CAMS\CapturedFiles\... session.
- Post-capture processing will MOVE the capture session to D:\CAMS\SubmissionFiles
- An initial attempt to perform the Upload logic on the new capture session(s) will be performed using the D:\CAMS\SubmissionFiles session(s).
- Later in the afternoon, a script will run (similar to the archiving scripts now) that MOVES all the non-transmitted SubmissionFiles sessions to the E:\CAMS\SubmissionFiles location.
  (This is where I haven't figured out all the logic yet. I haven't decided to migrate the new data to the tier-2 archive or to replicate it and leave the original copy on Tier-0 for a day or two. The thinking is to achieve better data permanency)

LaunchCapture has 5 phases. For each capture session, there are 4 post-capture steps (2-4):

1. Capture
2. Validation
3. AutoCal
4. Detect and ApplyCal
5. Move to SubmissionFiles
6. Perform post-capture processing on all remaining CapturedFiles sessions.

Outside of LaunchCapture, there are other phases for maintaining the system. Those are described in the subsequent sections.

UPLOADING

Uploading is accomplished using the "Cams2_Upload_Queue" task. Usually it runs in the morning after there has been enough time for all the post capture processing to finish for all the boards. Cams2_Upload_Queue first calls ZipCams2FromFTP.bat, which compares the cams2_queue\Transmitted dir against all the SubmissionFiles dirs. If there are any missing, then it creates a new ZIP file from the FTP folder in the SubmissionFiles session and moves it and its matching md5.txt file to the queue dir. ZipCams2FromFTP.bat also moves the weather forecast file into the SubmissionFiles\...\FTP dir. When all the SubmissionFiles sessions that have not yet been transmitted are in the queue dir, then the upload_queue.bat script continues by performing the robust upload logic by calling ftp_upload_robust_seti.bat for each zip file.

ARCHIVING

At about 3pm each day, the "Cams2 Archiving" scheduled task runs. It calls "archive_allbut_n_sessions.bat" to perform the logic. This script checks the INI file for the different Maximum days to keep for SubmissionFiles sessions, Transmitted zip files, and Cal files. Typically, CAL files are kept for about a year so that we can monitor trends in pointing drift. Transmitted zip files are kept for about 60 days in case they need to be retransmitted they are easy enough to locate. SubmissionFiles are the largest ones. Ideally, we'd keep 30 days of these. However, because of the size of the CAMS hard drive, it may be as few as 3-5 days. In the cams_Archive, we'd like to keep about 2 year's worth of these in the archive. Ideally, this archiving system was designed to archive to an external drive that is separate from the CAMS drive. If the archive drive is removeable, it will be an easy task to replace a full archive drive with a new empty one.

Cams2 Reboot PC

At about 4pm each day, the "Cams2 Reboot PC" scheduled task runs. It calls the Windows SHUTDOWN.EXE program to reboot the computer. We have found that rebooting the PC at least once per day improves the reliability of the network interface.

LAUNCHCAPTURE_KILL

This is a script that typically runs about 10 minutes before the LaunchCapture scheduled tasks. It only takes a minute or two to run.
LaunchCapture_KILL also calls the code to collect the forecast file and save it into the cams2_queue\Weather directory so that it can be copied into the SubmissionFiles\...\FTP dir later.

Cams2_Data_Plan_Reboot
After a reboot, whenever the user logs in (or unattended login happens), 10 minutes after the login this task will run the "Cams2_data_plan_reboot.bat" script. This script resets all the fields pertaining to a reboot so that it resets the offsets in the BytesThisBillingCycle.txt file. We have to do this because the only tracking method we have is to query the Network Bytes using NETSTAT -e to see if we've overrun the data plan limits. The network stats reset to 0 upon reboot. This script also updates the "reboot date string" to whatever is returned by NET STATISTICS WORK.

Cams2_Data_Plan_Reset
When the day specified in the INI file for BillingDate occurs, the Cams2_data_plan_reset.bat script is launched. Its job is to change the "data plan month" to the current month, reset the data plan for that month to 0,0,0; and reset the offsets to 0,0,0.

Cams2_data_plan_update
The cams2_data_plan_update.bat script updates the values in the cams2_queue\BytesThisBillingCycle.txt file. It also appends to the "data_plan_update.log" so that there is a track record of the changes to the data plan logging.

From: Andy Howell [mailto:camsflorida@gmail.com]
Sent: Sunday, February 3, 2019 3:07 PM
To: mogulskier <mogulskier@gmail.com>
Subject: Re: windows defender

Dave,

Thanks, I will add a share.

I've uninstalled Avast antivirus and re-enabled Windows Defender. I made the exclusions you suggested.

The 16TB Drobo is DAS (not NAS). I have a feeling the issue is read/write speed. What thoughts do you have about the required read/write speed?

Best, Andy

On Sun, Feb 3, 2019 at 5:30 PM mogulskier <mogulskier@gmail.com> wrote:

> Add a share.
>
> - For **Name**, enter: CAMS-FL/CAMS-<first camera of the site>-<name of the site in the camerasites.txt file>-Dave Samuels
> - For **Description**, enter whatever you want or leave blank.
> - For **Full access**, check the box.
> - Click **Next**.
> - Check the "**Validity**" checkbox.
> - Click **Next**.
> - Check the "**Allow direct link**" checkbox.
> - Check the "**Allow users**" checkbox.

- In the big square box, enter: **mogulskier@hotmail.com.**
- Click **Accept**.

You can add multiple registered users to a single share or you should create a separate share for each user or group odf users.

Go Patriots!!!

Dave

From: Andy Howell [mailto:camsflorida@gmail.com]
Sent: Sunday, February 3, 2019 2:23 PM
To: mogulskier <mogulskier@gmail.com>
Subject: Re: windows defender

Dave,

I will make these changes . . . On a different note, I am installing DWService on the 2nd CAMS computer. Currently, it's on my home network. But it's designated for another site.

How do I configure so that you can log in?

Best, Andy

On Sun, Feb 3, 2019 at 4:04 PM mogulskier <mogulskier@gmail.com> wrote:

> Andy,

I'm wondering if your system is running slow because of windows defender? Typically, removable drives are scanned and trusted less than internal drives.

Try adding exclusions for all the base cams dirs., such as E:\Cams2_queue, E:\Cams2_board0, and E:\cams_Archive.

Also, if you have a newer version, it also supports excluding processes. If you have that option, you might also exclude both WinSCP.exe and WinSCP.com.

I just discovered that this is an issue while monitoring a new system in New Zealand and it made me think of your system (for the others CC'd on this message, Andy has a 16 TB Drobo NAS device that is connected to the CAMS computer using a USB3 interface)…

You can monitor too by doing this:

- Start **Task Manager**
- Click on the **Performance** tab
- At the bottom, click on **Open Resource Monitor** link
- In the Resource Monitor, click on the **Overview** tab.
- In the Overview tab, **collapse CPU** and **expand Disk**.
- Then **sort** it descending by the **Total** column
- When you're doing certain activities, you'll notice in the Image column that sometimes the process **MsMpEng.exe** appears. That's Windows Defender. If that happens, say during post-capture processing, autocal, detection, or upload, you might need to adjust windows defender some more.

You might feel that this is risky. Here are my answers to that:

- I've always had a problem with on-access scanning and I think it's stupid if your system is scanned one time and then you subsequently scan all the programs that are loaded onto your computer from the network or USB drives. After that, performing on-access scans for, say .txt files, is really a waste of processing time. Also, I think the way we do virus scanning is stupid (looking for byte patterns in files). If we instead used encryption and hashing technology, it would be much quicker and reliable and it wouldn't require virus database updates. For files, such as office documents and PDF files, we would need to do it differently, since those tools can have executable code in them.
- Having said that, first, I agree that it is risky if a virus gets into any of the FTP_*.exe or other programs, such as du.exe, process.exe, FindAndReplace.exe, etc. and you are right to always use caution and stay aware of the risks.
- However, since these programs are placed on your system through the a network interface, they are always scanned. We are only disabling on-access scanning for files that are already on the system but we leave all other scanning enabled.

- There is no reason to scan the files created during capture or post-capture processing, since those are created by already-scanned programs on your computer. These are essentially data files and they don't include executable code in them.
- .BAT files are essentially text files. They can be infected by someone adding lines that create code and run it, but these files would also be scanned whenever these scripts are copied to your computer across the network.
- You might be also worried about transmitting viruses when we upload to SETI. Well, since the .bin files are created on your system, there is no reason to scan them for on-access, such as during autocal or detection. The zip files they are in will still be scanned when they are transmitted.
- Therefore, I feel that disabling on-access scanning for the cams directories, while leaving it enabled for other things, will improve performance (possibly by double) and still keep the system safe.

I hope this helps. Pass this along to others as you see fit.

Dave