Data capture jobs are used to identify data on your network and stream it to Transfer Appliance. The data capture options are:

- Workstation capture: To perform workstation capture, you install a Capture Utility on a Windows (/transfer-appliance/docs/2.0/capturing-data-windows) or Linux (/transfer-appliance/docs/2.0/capturing-data-linux) workstation. Workstation captures typically deliver the best performance.

- NFS export: Exporting an NFS Share (/transfer-appliance/docs/2.0/exporting-nfs-share) lets you configure Transfer Appliance to export directories via NFS. This allows you to mount the exports on any NFS client workstation or server, and copy the data to the export. Use this method if you cannot install the Capture Utility on the workstation, or if the workstation is not an NFS server. After copying, the Transfer Appliance processes the NFS share to compress and encrypt the copied data.

- HDFS capture: Capturing data from HDFS (/transfer-appliance/docs/2.0/capturing-data-hdfs-nfs-share) this capture method is similar to NFS export, in that you configure Transfer Appliance to export directories via NFS. This allows you to mount the exports on HDFS, and then copy data from HDFS to the export. After copying, the Transfer Appliance processes the NFS share to compress and encrypt the copied data.

- NFS capture: Performing an NFS capture (/transfer-appliance/docs/2.0/capturing-data-appliance) lets you connect directly to an NFS share from Transfer Appliance, which means you don't need a separate workstation to run the Capture Utility. This is a simple capture method, provided your network is secure.

The total available capacity for capturing data is affected by the type of capture you choose to use.

The following table summarizes the available capacity depending on the capture type:

Capture type	TA100	TA480
Workstation capture	81 TB	423 TB
NFS Capture	81 TB	423 TB

NFS export and HDFS capture	74 TB	400 TB

You can run multiple capture jobs simultaneously. For example, you can run multiple NFS and workstation capture jobs simultaneously. The number of parallel capture jobs is limited only by your system resources and network bandwidth capacities. It is recommended that you run multiple capture jobs at once in order to transfer data faster.

The Capture Utility automatically spawns up to 8 parallel capture tasks for each capture job, with each task handling up to 1 terabyte (TB) of data. This helps optimize performance and bandwidth utilization. As each capture task completes, a new one is created, until all targeted data has been captured.

- Assign jobs unique, meaningful names. Job names are used to identify each capture job, and the files it contains, for the rest of the data migration project.

    For example, capturing the file:

    `e:\sourcedatafolder\data1\file_001`

    with the Windows Capture Utility command:

    `tacapture.exe this-job e:\sourcedatafolder\data1`

    creates the file:

    `gs://<bucket_name>/this-job/e/sourcedatafolder/data1/file_001`

    in the Cloud Storage staging location when your data is uploaded into Cloud Platform.

- If you plan to export NFS shares from Transfer Appliance for data capture, create one NFS share for each planned job and initiate NFS share processing after the data is copied onto the NFS share. Make sure there is 7 TB total space available to process NFS shares on Transfer Appliance.

- Plan to allocate time to verify the captured data before shipping Transfer Appliance back to Google. Allocate time to process NFS shares if you perform an NFS share capture. The time required to verify and process the data depends on the amount of data collected and the deduplication ratio. The approximate time required for verification is displayed in Transfer Appliance web interface when you start the Prepare for Shipping (/transfer-appliance/docs/2.0/shipping-appliance) operation.

By default, a data capture job dynamically chooses a port range to use. Each data capture task in the job requires its own data streaming port and chooses one from the range between the starting port number and (port number) + (data capture tasks) - 1. For example, if the data capture job starts at port 50555 and uses the default of 8 for data capture tasks, the data capture tasks would use ports 50555-50562.

If a capture job fails, it can be restarted. Capture jobs checkpoint their progress, so a restarted capture job resumes from the last known good point to capture remaining data. For more information, see Retrying Failed Jobs (/transfer-appliance/docs/2.0/retrying-failed-jobs).

Transfer Appliance Capture Utility calculates the CRC32C hash of each file while capturing it. This hash value is used for integrity checking during the data rehydration process. For more information, see Verifying rehydrated data (/transfer-appliance/docs/2.0/rehydrating-data#verify-rehydrated-data).

To prepare for your transfer, see Before you order an Appliance (/transfer-appliance/docs/2.0/preparing-data-transfer).