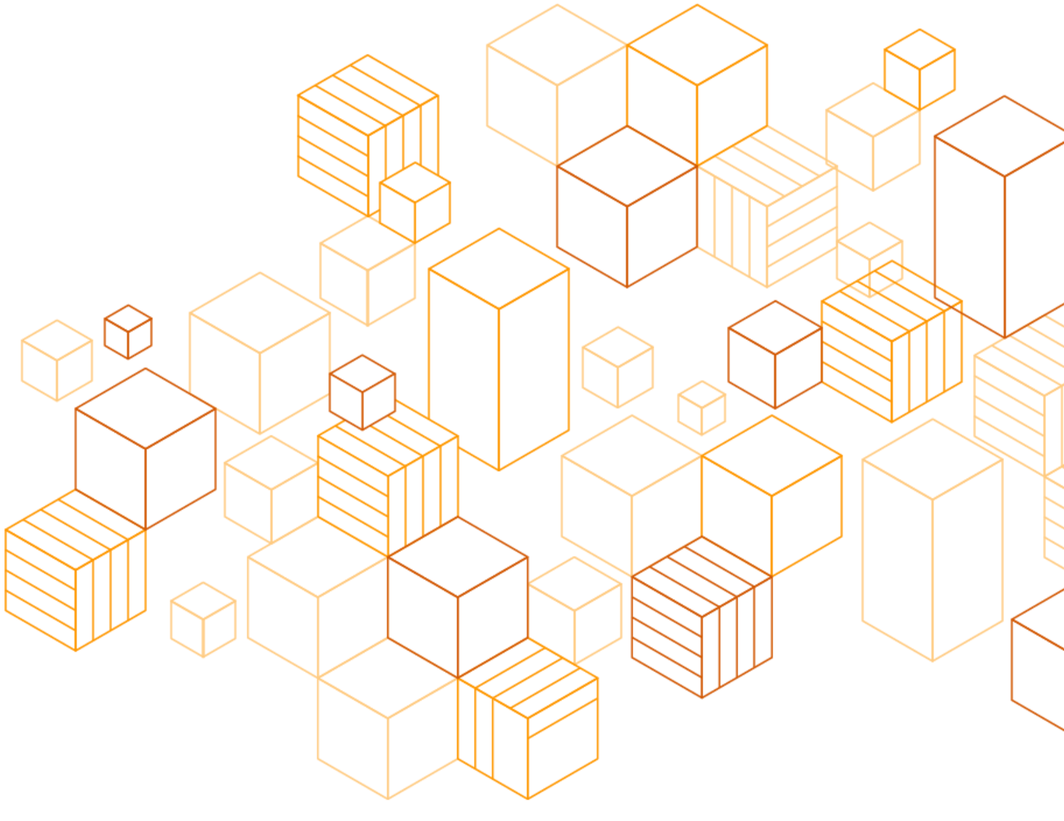
AWS Snowball Edge Data Migration

Guide

April 2020



Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents current AWS product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers or licensors. AWS products or services are provided “as is” without warranties, representations, or conditions of any kind, whether express or implied. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

Contents

About this Guide

This document provides best practices for AWS Snowball Edge bulk data transfers between on-premises data centers and the AWS Cloud.

This data migration guide applies only to Snowball Edge devices. It does not apply to AWS Snowball devices.

Considerations

- Transfer performance to the Snowball Edge device depends upon multiple factors including: the size of files being transferred, read performance of the source data storage system, tools used for the data transfer, network link speed, network utilization, network latency, and the number of network hops between the data transfer client and the Snowball Edge unit. Consider these aspects when planning for a Snowball Edge job.
- Use AWS Command Line Interface (AWS CLI) with Amazon S3 to improve transfer performance by employing multiple sessions and parallelizing operations when transferring large volumes of data.
- To accurately predict the total time required to migrate large datasets to Amazon S3 using Snowball Edge, you must first run benchmarks and measure performance in your own environment. The time to import a job from the Snowball Edge device into Amazon S3 is comparable to the time required to transfer data from on-premises storage systems into the Snowball Edge unit.
- Consider the File interface (NFS) when you need to preserve the metadata of files and the performance requirements are low. Use the Amazon S3 interface for performance demanding data transfers. The maximum file size that can be transferred to Snowball Edge using the File interface is 150 GB.
- The File interface on Snowball Edge implements a subset of the standard NFS specifications. The File service runs on a separate IP address that must be created using the `snowballedge` command line tool before enabling it. See Using the File Interface for the AWS Snowball Edge for more information.
- After the import job completes, you can review the Snowball job reports from the AWS Management Console. See Verification for more information.

Before You Order

AWS Snowball Edge is a Region-specific service. Verify that the service is available in your Region before you begin planning your migration. See Regional Table to verify availability.

Ensure that your location and bucket are within the same Region or country. The AWS Management Console does not allow device orders if the bucket and Region are not in the same location. Limitations to shipping outside of a Region’s country borders are outlined in the AWS Snowball Developer Guide. International shipments to locations outside of the Region will be supported through the AWS Management Console for whitelisted customers for a select set of locations from specific Regions (like US to Mexico). Discuss the target destination, costs, and timing for your requests with your account team.

As part of the order process, you are asked to create an AWS Identity and Access Management (IAM) role and AWS Key Management Service (AWS KMS) key. The AWS KMS key encrypts the data during transit and at rest on the Snowball Edge device. For more information on creating IAM roles and KMS keys, see the AWS Snowball Developer Guide.

You can use the following steps to encrypt the imported data in the Amazon S3 bucket using the Server-side Encryption Key Management System (SSE-KMS) encryption:

1. In the AWS Management Console, enable **Default Encryption** on the Amazon S3 Bucket into which the data will be imported. Select **AWS KMS** as the encryption method, and then select a specific KMS key. This is the key used to encrypt objects imported into the bucket.
2. After the Snowball Edge job is created, but before the data is imported, you must add a statement to the existing **IAM Role for an Import Job** policy. This role is created when the Snowball Edge job is created. The statement must be similar to the following:

```
{
     "Effect": "Allow",
     "Action": [
       "kms:GenerateDataKey"
     ],
     "Resource": "arn:aws:kms:us-west-
2:123456789012:key/abc123a1-abcd-1234-efgh-111111111111"
}
```

This example statement grants the necessary permission on the SSE-KMS. For more information, refer to Creating an IAM Role for Snowball.

Selecting Snowball Edge Services

Snowball Edge devices are shipped with Amazon S3 endpoint and File service capabilities.

If you plan to run additional services, such as AWS Lambda on AWS IoT Greengrass, or Amazon Elastic Compute Cloud (Amazon EC2) instances, additional steps must be completed before you order.

Refer to the following documentation for more information:

- Amazon EC2 on Snowball Edge
- AWS IoT Greengrass and Lambda Functions on Snowball Edge

Preparing the Client Workstation

A client workstation is used to unlock and manage the Snowball Edge and to transfer data to the device. The client workstation must run a 64-bit OS — Linux (Ubuntu version 12 or later, and RHEL version 6 or later), Windows 7 or later, or macOS 10.10 or later. The following packages must be installed on the client workstation before proceeding:

- AWS CLI version (1.16.14 or earlier)
- Snowball Edge client

Network Connections and Ports

There are three networking options available on the Snowball Edge:

- 10 Gb/s Ethernet (RJ45) (operates at either 1 Gb/s or 10 Gb/s)
- 25 Gb/s with SFP connection
- 40 Gb/s with QSFP connection

The most common connection method is the RJ45 interface with a CAT-5/6 cable. The network connection between the client workstation and the data source must have low network latency to maximize performance. *Figure 1* shows a simple configuration.

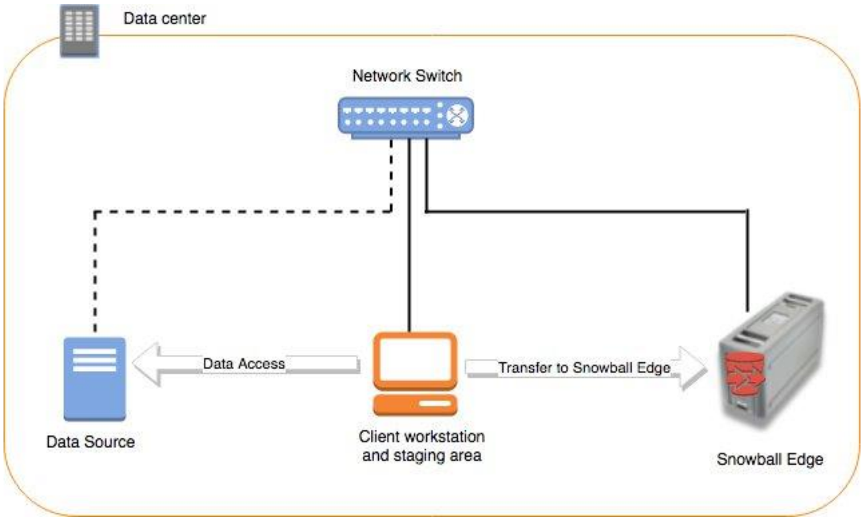

Data center
Network Switch
Data Access
Transfer to Snowball Edge
Data Source
Client workstation and staging area
Snowball Edge

Figure 1: Typical Snowball Edge connectivity diagram

The SFP and QSFP options can be used for higher throughput, but they require additional hardware and cables to connect. The recommended SFPs and cables as listed in the AWS Snowball Edge Specifications.

To transfer data using the Amazon S3 CLI, the following TCP ports must be open between the client workstation and the Snowball Edge device:

- 9091 - Activation port on Snowball Edge
- 8080 - The HTTP endpoint for Amazon S3 on Snowball Edge
- 8443 - The HTTPS endpoint for Amazon S3 on Snowball Edge
- 22 - SSH access to support diagnostics

For data transfer using the File interface (NFS), the following ports must be open:

- 111 (TCP and UDP)
- 2049 (TCP and UDP)
- 20048 (TCP and UDP)

Physical Installation

Perform a physical inspection before installing the Snowball Edge in your data center. Look for any tampering or damage during shipment. If there is evidence of tampering, contact AWS Support immediately. AWS will not import data from the Snowball Edge device into the Amazon S3 bucket if there is evidence of tampering.

If the Snowball Edge device will be installed in an enclosed rack, leave room for air circulation to the front of the unit. When the Snowball Edge is powered on, you must leave the front door open. Snowball Edge is built as a ruggedized unit and does not have the ability to be mounted using rack mounting rails, but can sit on rack shelves.

Electrical and Network Connection

The Snowball Edge uses a single power source and is rated for 400 W power consumption. Refer to the AWS Snowball Edge Specifications for more information. User the following process to connect the Snowball Edge to power and to the network:

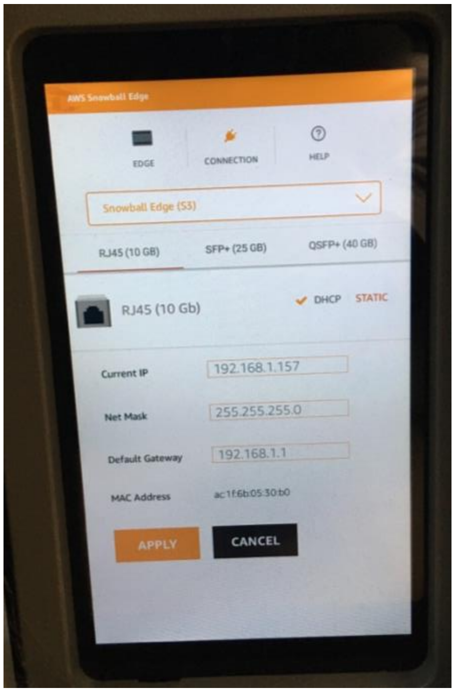
3. Access the touch screen, the power port, and the network port by opening the front and back sliding doors on the Snowball Edge.
4. Open the door on the top of the Snowball Edge and remove the provided power cable from the cable nook. The network cable is not supplied.
5. Plug in the power and network cables.

Powering the Device

Press the power button located above the LCD display. Once powered on, the device takes 15-20 minutes to complete internal validation checks. While the device is validating, the LCD display shows a short instructional video.

Configuring the Snowball Edge IP Address

1. Select Connection on the touch screen panel, and then select DHCP or Static to specify which IP address assignment method to use.
2. If you are using the Static IP addressing scheme, enter the IP address, Net Mask, and Default Gateway. If you are using DHCP, the IP address is automatically assigned by the DHCP server on your network.
3. Apply the changes.



RJ45 (10 Gb)

Current IP 192.168.1.157

Net Mask 255.255.255.0

Default Gateway 192.168.1.1

MAC Address ac:1f:6b:05:30:b0

Figure 2: Snowball Edge front panel display with IP address configuration menu

Checking Network Connectivity

Record the IP address of the Snowball Edge. If you selected the DHCP option, the IP address is automatically assigned by your network's DHCP server. From the client workstation, ping the Snowball Edge IP address.

If your ping to the IP address returns `no connect`, run the following steps to diagnose the error:

1. Check the link status at the back of Snowball Edge. The link LED should be on, signaling that the network connection between the switch and the Snowball Edge is established.
2. If the link LED is not illuminated, check the cable and switch port status. On a managed switch, you are able to check the port speed, port status, and duplex settings using the switch user interface.
3. Verify the link status of the network interface on the client workstation.
4. Confirm that the client workstation and Snowball Edge are on the same IP Network in terms of connectivity and IP addressing scheme.
5. If your network uses VLANs, the client and Snowball Edge must be on the same VLAN.

6. Check the firewalls in your network. The Snowball Edge IP address must be whitelisted on the client workstation.
7. Unplug and re-connect the RJ45 cable on the Snowball Edge device. Ping again.

If the ping attempts continue to fail, follow these steps:

1. Power off the Snowball Edge.
2. Unplug the power and network cables.
3. Wait 10 minutes.
4. Re-connect power and network cables.
5. Power on the Snowball Edge.
6. Wait 20 minutes and test by pinging again.
7. Check if a direct connection works:
    a. Use an Ethernet cable to connect the Snowball Edge directly to the client workstation.

    b. Configure the workstation with an IP address in the same address range as the Snowball Edge IP address and netmask.

    c. Try pinging the Snowball Edge IP address again. This rules out any issues with the Snowball Edge NIC and connectivity.

    d. If the direct connection in step b is successful, engage your Networking team to troubleshoot connectivity issues.

Troubleshooting Transfer Issues

1. Verify that you are not experiencing IP address conflicts. Confirm the MAC address in your client's arp table matches the MAC address of the Snowball Edge LCD screen. You can run `arp -a` from the terminal window to derive the IP address and MAC address mappings, as shown below:

```
client_workstation/>  arp -a
? (192.XXX.XX.XXX) at x:xx:xx:xx:xx:xx on en0 permanent [ethernet]
? (192.XXX.XX.XX) at xx:xx:xx:xx:xx:xx on en0 ifscope permanent
[ethernet]
client_workstation/>
```

2. If the transfer rates are low, find the source of the bottleneck. Possible sources could be:

    e. Multiple hops from transfer workstation or data source to the Snowball Edge. To troubleshoot, `run >traceroute <snowball_ip>` to identify the number of hops.

    f. If possible, check the Round- Trip Transit times (RTT) of packets using `>tcpdump`. A longer latency results in a slower transfer.

    g. Check the port speeds for Snowball Edge, the client workstation, and the data source from the switch management interface.

    h. Check the duplex settings on the switch port using switch management commands (this is available only on managed switches).

    i. Perform a simple file transfer operation to another host in the network to better understand read performance and network throughputs.

    j. Test to see if the following ports are accessible from the client workstation:

```
telnet <snowball_ip> <port_number> where the port number is; 9091
(Activation port),  22 (ssh) and  8080 (http endpoint for s3)
```

Unlocking Snowball Edge

1. Using your web browser, log in to the AWS Management Console and select the Snowball service.

2. Select the Snowball Edge job ID that corresponds to your device. The job ID is a unique 39-character label that identifies your job and can be found at the bottom of the shipping label on the E Ink display of the Snowball Edge unit. The job ID is also included in the name of the job's manifest file.

3. Download the manifest file and record the unlock-code.

4. From the command line on your client workstation where the Snowball Edge client software is installed, configure Snowball Edge credentials by running the following code:

```
>snowballedge configure --profile <profile_name>
client_workstation/>pwd
/home/user1
client_workstation/>ls
```

```
J1aabbccdde-1234-1234-c123456789abc_manifest.bin
client_workstation/> snowballedge configure --profile sbe1
Configuration will stored at
/home/user1/.aws/snowball/config/snowball-edge.config
Snowball Edge Manifest Path: /home/user1/J1aabbccdde-1234-1234-
c123456789abc_manifest.bin
Unlock Code: 3fc80-e7565-7d25d-41926-ac27a
Default Endpoint: https://XXX.XXX.XX.XXX
client_workstation/>
```

This configuration command prompts the Snowball Edge endpoint, the manifest file location, and the 29-character unlock code.

- The manifest file location must be the complete path to your file. For example: `/home/user1/ABC1234567891234-12a4-567b-89c0-de1234f56Gg_manifest.bin`
- The end-point must be a full URL, with an https prefix. For example: https://XXX.XXX.XX.XX, where XXX.XXX.XX.XX is your Snowball Edge IP address.

Unlock the Snowball Edge by running the following command:

```
>snowballedge unlock-device  --profile <profile_name>
client_workstation/> snowballedge unlock-device --profile sbe1
Your Snowball Edge device is unlocking. You may determine the
unlock state of your device using the describe-device command. Your
Snowball Edge device will be available for use when it is in the
UNLOCKED state.
client_workstation/>
```

Verify the Snowball Edge lock status by running the following code:

```
>snowballedge describe-device --profile <profile_name>
client_workstation/> snowballedge describe-device --profile sbe1
{
  "DeviceId" : "JID14a49650-56d6-47ba-xxxxxxxxx",
  "UnlockStatus" : {
    "State" : "UNLOCKING"
  },
  "PhysicalNetworkInterfaces" : [ {
    "PhysicalConnectorType" : "RJ45",
    "IpAddressAssignment" : "DHCP",
    "IpAddress" : "XXX.XXX.XX.XXX",
    "Netmask" : "255.255.255.0",
    "DefaultGateway" : "XXX.XXX.XX.X",
    "MacAddress" : "xx:xx:xx:xx:xx:xx"
  }, {
    "PhysicalConnectorType" : "SFP_PLUS",
    "IpAddressAssignment" : "STATIC",
    "IpAddress" : "0.0.0.0",
    "Netmask" : "0.0.0.0",
    "DefaultGateway" : "XXX.XXX.XX.X",
    "MacAddress" : "xx:xx:xx:xx:xx:xx"
  }, {
    "PhysicalConnectorType" : "QSFP",
    "IpAddressAssignment" : "STATIC",
    "IpAddress" : "0.0.0.0",
    "Netmask" : "0.0.0.0",
    "DefaultGateway" : "XXX.XXX.XX.X",
    "MacAddress" : "xx:xx:xx:xx:xx:xx"
  } ]
}
client_workstation/> snowballedge describe-device --profile sbe1
{
  "DeviceId" : "JID14a49650-56d6-47ba-xxxxxxxxx",
  "UnlockStatus" : {
    "State" : "UNLOCKED"
  },
  "ActiveNetworkInterface" : {
    "IpAddress" : "XXX.XXX.XX.XXX"
  },
  "PhysicalNetworkInterfaces" : [ {
    "PhysicalNetworkInterfaceId" : "s.ni-xxxxxxx",
    "PhysicalConnectorType" : "RJ45",
    "IpAddressAssignment" : "DHCP",
    "IpAddress" : "XXX.XXX.XX.XXX",
    "Netmask" : "255.255.255.0",
    "DefaultGateway" : "XXX.XXX.XX.X",
    "MacAddress" : "xx:xx:xx:xx:xx:xx"
  }, {
    "PhysicalNetworkInterfaceId" : "s.ni-xxxxxxx",
    "PhysicalConnectorType" : "QSFP",
    "IpAddressAssignment" : "STATIC",
    "IpAddress" : "0.0.0.0",
    "Netmask" : "0.0.0.0",
    "DefaultGateway" : "XXX.XXX.XX.X",
    "MacAddress" : "xx:xx:xx:xx:xx:xx"
  }, {
    "PhysicalNetworkInterfaceId" : "s.ni-xxxxxxx",
    "PhysicalConnectorType" : "SFP_PLUS",
    "IpAddressAssignment" : "STATIC",
    "IpAddress" : "0.0.0.0",
    "Netmask" : "0.0.0.0",
    "DefaultGateway" : "XXX.XXX.XX.X",
    "MacAddress" : "xx:xx:xx:xx:xx:xx"
    } ]
 }
client_workstation/>
```

The status can be also verified from the front-panel display:

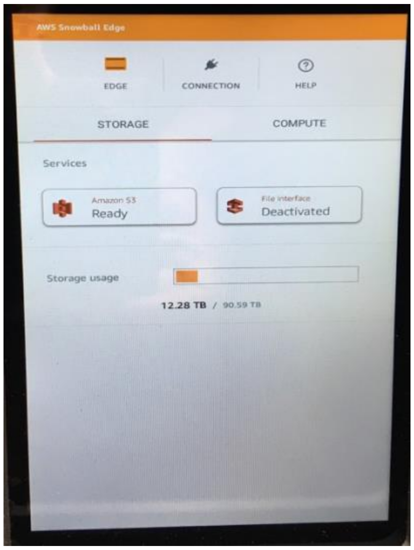


Figure 3: Unlocked Snowball Edge front panel display

Each Snowball Edge device only needs to be unlocked once, regardless of if you have multiple client workstations. If the Snowball Edge is power cycled, it will lock. Repeat the steps above to unlock the device again.

Obtaining Snowball Edge Access Keys

In order to run Amazon S3 commands from the AWS CLI, you must retrieve and store access credentials on your client workstation.

1. List the available access keys on the Snowball Edge device, using the `>snowballedge list-access-keys` command:

```
client_workstation> snowballedge list-access-keys --profile sbe1
 {
  "AccessKeyIds" : [ "XXXXXXXXXXXXXXXXXXXX" ]
 }
client_workstation/>
```

2. Use the `>snowballedge get-secret-access-key` command to get the corresponding secret key:

```
client_workstation> snowballedge get-secret-access-key --access-
key-id XXXXXXXXXXXXXXXXXXXX --profile sbe1
[snowballEdge]
aws_access_key_id = XXXXXXXXXXXXXXXXXXXX
aws_secret_access_key = /+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
[user1@oecent]
```

3. Use the `aws configure` command to store the keys so they don't need to be specified with each Amazon S3 command. The keys are stored in the `~/.aws/credentials` file on the client workstation.

```
client_workstation> aws configure --profile sbe1
AWS Access Key ID [None]:XXXXXXXXXXXXXXXXXXXX
AWS Secret Access Key
[None]:/+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Default region name [None]: snow
Default output format [json]:
client_workstation/>
```

Performing Your First Copy Operation

Use the following commands to validate Amazon S3 functionality to the Snowball Edge device. These operations are performed using the http endpoint on the Snowball Edge. Using the https endpoint requires additional steps to obtain the proper certificates. Refer to the AWS Snowball Developer Guide for more information.

1. List the Amazon S3 buckets on your Snowball Edge. The buckets were configured when you created the Snowball job.

```
client_workstation>aws s3 ls --endpoint http://xxx.xxx.xx.xxx:8080
--profile sbe1
2009-10-12 13:50:30 sbeingest2018
client_workstation/>
```

2. Perform a simple file copy using the `s3 cp` command:

```
client_workstation> aws s3 cp --recursive /data/smallfile_out_1
s3://sbeingest2018 --endpoint http://xxx.xxx.xx.xx:8080 --profile
sbe1
upload:
../../../data/smallfile_out_1/file_srcdir/oecent.localhost/thrd_00/
d_000/_oecent.localhost_00_12_ to
s3://sbeingest2018/file_srcdir/oecent.localhost/thrd_00/d_000/_oece
nt.localhost_00_12_
client_workstation/>
```

3. List the files you just copied using the `s3 ls` command:

```
client_workstation> aws s3 ls --recursive s3://sbeingest2018/ --
endpoint http://xxx.xxx.xx.xx:8080 --profile sbe1
2018-09-19 09:35:29       2001 TEST123.txt
2018-10-04 09:11:48    1048576
file_srcdir/oecent.localhost/thrd_00/d_000/_oecent.localhost_00_10_
2018-10-04 09:11:48    1048576
file_srcdir/oecent.localhost/thrd_00/d_000/_oecent.localhost_00_11_
2018-10-04 09:11:47    1048576
file_srcdir/oecent.localhost/thrd_00/d_000/_oecent.localhost_00_12_
2018-10-04 09:11:48    1048576
file_srcdir/oecent.localhost/thrd_00/d_000/_oecent.localhost_00_13_
client_workstation/>
```

4. Use the `s3 sync` command to copy a directory tree with all its content to the Snowball Edge:

```
client_workstation> aws s3 sync /var/log
s3://sbeingest2018/datasync --endpoint http://xxx.xxx.xx.xx:8080 --
profile sbe1
client_workstation/>
```

Logging Output from Transfer Operations

Run the following command to generate a list of files as they are being copied:

```
client_workstation> aws s3 sync /data s3:/sbingest2018/data --
endpoint http://xxx.xxx.xx.xxx:8080 --no-progress --profile
sbe1 >/tmp/transfer_log_1 2>&1
```

The above command redirects the output and error of the AWS CLI to the file you specified — in this case, `transfer_log_1`. To view the output, open another console window and tail the file:

```
client_workstation> tail -f /tmp/transfer_log_1
```

Transferring Large Number of Small Files

Batching may be used to improve performance while transferring a large number of small files (files less than 1 MB). Snowball Edge does not support automated batching of small files. To manually batch and then transfer files to Snowball Edge you may run the following command:

```
client_workstation>tar -cf - /Logs/April | aws s3 cp -
s3://mybucket/batch01.tar
--metadata snowball-auto-extract=true --endpoint
http://xxx.x.x.x:8080 --profile sbe1
client_workstation/>
```

When batching small files, keep the following in mind:

- The recommended number of files per batch is 10,000, with a maximum batch size of 100 GB
- Batched files are auto-extracted to the Amazon S3 bucket during import
- Batches larger than 100 GB will not be auto-extracted when imported to Amazon S3.
- Supported archive formats are tgz, tar, and ZIP.

Benchmarking Transfer Performance

To benchmark transfer performance, take a sample set of files matching your workload profile, and perform multiple iterations of the `s3 sync` or `s3 cp` commands. Record the throughput characteristics each time. The transfer performance can be improved by following these guidelines:

- Perform multiple copy operations in parallel
- Copy from multiple client workstations
- Transfer directories, not files

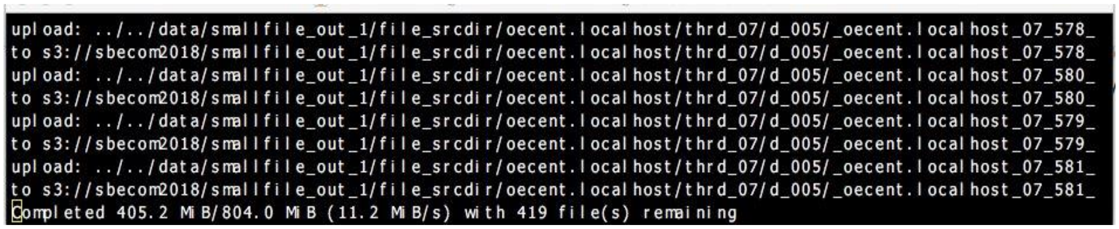
- Don't perform other operations on source files during the transfer
- Reduce local network use

Review the AWS CLI S3 Configuration documentation for more information.

Monitoring

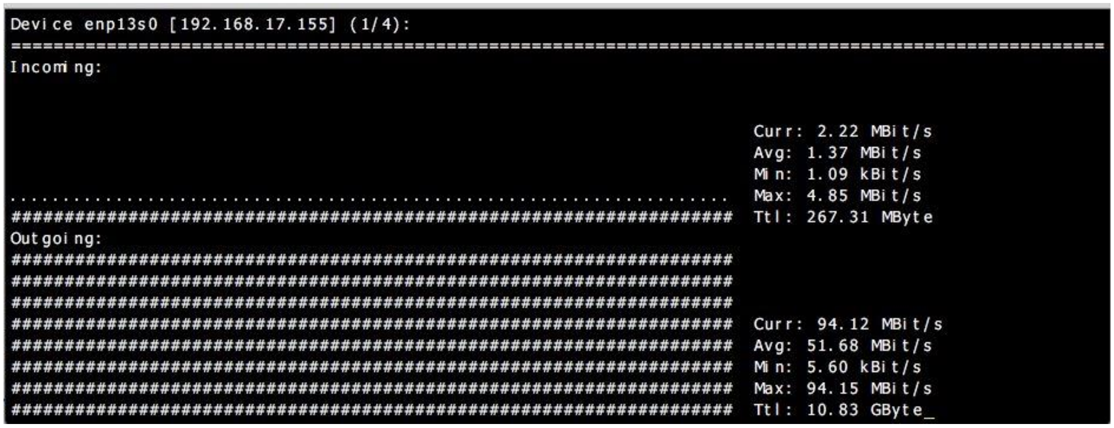
You can use `nload` to monitor transfer speed to the Snowball Edge device. If you find an issue during your benchmarking, isolate the problem areas before beginning the full transfer. The following example demonstrates running an `s3 sync` operation and monitoring the network speed using `nload`.

Sample output from the `s3 sync` command:



```
upload: ../../data/smallfile_out_1/file_srcdir/oecent.localhost/thrd_07/d_005/_oecent.localhost_07_578_
to s3://sbecom2018/smallfile_out_1/file_srcdir/oecent.localhost/thrd_07/d_005/_oecent.localhost_07_578_
upload: ../../data/smallfile_out_1/file_srcdir/oecent.localhost/thrd_07/d_005/_oecent.localhost_07_580_
to s3://sbecom2018/smallfile_out_1/file_srcdir/oecent.localhost/thrd_07/d_005/_oecent.localhost_07_580_
upload: ../../data/smallfile_out_1/file_srcdir/oecent.localhost/thrd_07/d_005/_oecent.localhost_07_579_
to s3://sbecom2018/smallfile_out_1/file_srcdir/oecent.localhost/thrd_07/d_005/_oecent.localhost_07_579_
upload: ../../data/smallfile_out_1/file_srcdir/oecent.localhost/thrd_07/d_005/_oecent.localhost_07_581_
to s3://sbecom2018/smallfile_out_1/file_srcdir/oecent.localhost/thrd_07/d_005/_oecent.localhost_07_581_
Completed 405.2 MiB/804.0 MiB (11.2 MiB/s) with 419 file(s) remaining
```

`nload` output with current, average, and maximum outgoing transfer speeds:



```
Device enp13s0 [192.168.17.155] (1/4):
============================================================================================================
Incoming:



                                                                          Curr:  2.22 MBit/s
                                                                          Avg:  1.37 MBit/s
                                                                          Min:  1.09 kBit/s
......................................................................    Max:  4.85 MBit/s
######################################################################    Ttl:  267.31 MByte
Outgoing:
######################################################################
######################################################################
######################################################################
######################################################################    Curr:  94.12 MBit/s
######################################################################    Avg:  51.68 MBit/s
######################################################################    Min:  5.60 kBit/s
######################################################################    Max:  94.15 MBit/s
######################################################################    Ttl:  10.83 GByte_
```

You can use `python smallfiles` to create test data for transfer benchmarking.

Validating

The Snowball Edge device takes a number of steps to ensure that data is transferred properly to the device and from the device into AWS. However, it is best practice to verify the transfer independently. Consider the following when determining how to validate your overall data transfer:

- Before beginning the full data transfer, it is recommended that an inventory be taken at the data source. The inventory should include a list of every file to be transferred to the Snowball Edge, along with a checksum for each file. This allows for a full and complete comparison of data once it arrives at the Amazon S3 bucket. This step may take a significant amount of time and could significantly impact performance of the source data store, depending upon the number and size of files to be transferred.
- Checksums are automatically calculated on each file when transferred into the Snowball Edge. After the unit has been returned to AWS, file checksums are compared and validated while importing to the Amazon S3 bucket. Files that do not pass checksum validation are not imported. The **Download Failure logs** link on the Snowball job report provides a list of files that were not properly imported.
- Files that failed to import via Snowball will need to be copied manually from the client workstation to the Amazon S3 bucket in AWS.
- Take a subset of files and perform a checksum of the files at the source on-premises and in the destination Amazon S3 bucket. Verify that the checksums match.
- If you are performing checksums for a large number of files in Amazon S3, follow the procedure below to keep Amazon S3 data transfer within the Region and reduce egress charges:

  a. Launch an Amazon Linux instance on EC2 and verify that it has the appropriate role to access the Amazon S3 bucket.

  b. Install AWS CLI.

  c. Get files from the Amazon S3 bucket using `aws s3` commands.

  d. Perform checksums on the EC2 instance and compare them with your on-premises checksums.

Snowball Edge Size and Capacity Limitations

- The maximum size of a single file is 5 TB when AWS CLI is used for data transfer.
- The maximum size of a single file is 150 GB, when NFS is used for data transfer.

- The maximum usable capacity is 82 TB for a single storage-optimized Snowball Edge.
- The maximum usable capacity is 42 TB for a single compute-optimized Snowball Edge.

Additional Limitations

- Command line tools — `snowball` and `snowball-adapter`, will not work on Snowball Edge. These are legacy commands that applied to original Snowball devices no longer available as of April 7, 2020.
- The Snowball Edge is not designed to be field repairable. If it fails, it must be returned. You can obtain a replacement by placing a new job order.
- The manifest file associated with the job defaults to 360 days. However, this can be extended. To extend the file, you must contact AWS Support at the time of the job creation. The certificates installed on a Snowball Edge device have a life of only 120 days.
- When using the Amazon S3 interface to transfer data to the Snowball Edge, all file metadata (permissions, ownership, timestamps, etc.) will be replaced with the following defaults – only the file size and file name and path will remain unchanged:

```
-rw-rw-r-- 1 root root [filesize] Dec 31 1969 [path/filename]
```

- To retain file metadata, use the File interface for data transfer. When using NFS to transfer files to the Snowball Edge, metadata will be stored in the user-metadata portion of the corresponding Amazon S3 object.

Recommendations

- Enable Amazon Simple Notification Service (Amazon SNS) on the AWS Management Console to track your Snowball Edge job at every stage from order to job completion.
- While powered on, do not close the Snowball Edge doors. The cooling vents are blocked when the front doors on the unit are closed. The Snowball Edge will automatically shut down to prevent thermal damage.

- The primary Snowball Edge IP address can be changed directly from the display. Secure the device in a locked room or rack to prevent misuse.
- Protect the unlock-code and manifest file from unauthorized access.
- Ensure that link speeds are optimal (check switch or LED indications) before starting the transfer.
- Test and benchmark before setting the migration/transfer goals.
- When possible, run multiple AWS CLI command sessions in Amazon S3 from multiple client workstations to speed up the transfer. Running multiple operations in parallel improves overall system throughput.
- Check the Snowball import report for any issues, particularly errors such as files not imported.

Job Report and Import Logs

1. Sign in to the AWS Management Console, select your job or job part from the table, and expand the status pane. Three options appear: **Get job report**, **Download success log**, and **Download failure log**.
2. Choose the log you want to download.

Contributors

Contributors to this document include:

- Vinod Pulkayath, Sr. Storage Specialist SA
- Bryan Berezdivin, Global Storage Business Development Manager
- Brandon Skinner, Support Engineer, AWS Snow Family

Document Revisions

Date	Description
April 2020	Update to parameters and limitations.
November 2019	First publication