Data Integrity Check

Data backup is essential to business or organization and having a data backup plan is only as good as the integrity of the backup data. To ensure that this objective is met, AhsayCBS backup server and AhsayOBM/AhsayACB client provides an improved **Data Integrity Check (DIC)** feature where the end user can easily verify the integrity of the data stored on the backup destination(s) (i.e. AhsayCBS, Cloud storage, or Local storage) to ensure that the backup data is recoverable.

In backing up large or even small file(s), data corruption may still occur during a backup job or even in a post-backup job. Some of the possible causes are:

- Bad program exits (i.e. AhsayCBS/AhsayOBM/AhsayACB application terminated unexpectedly when an active backup job is in progress)
- Technical problems on the AhsayOBM/AhsayACB client machine (e.g. hardware failure, unexpected reboot, unexpected loss of power)
- Technical problems on the AhsayCBS backup server (e.g. hardware failure, unexpected reboot, unexpected loss of power, storage issues, human error)
- Technical problems on the cloud storage service

Since data corruption is always a possibility, the solution is to **identify** and then **remove corrupted files from the backup destination(s)**. Identifying and removing corrupted files from the backup destination(s) is mission critical as it measures the **integrity of the backup data** and its restorability.

The primary role of the Data Integrity Check is to identify and remove corrupted files from the backup destination(s). This will allow the next backup job to have an opportunity to back up these files again. **However, corrupted files which are located in the retention area will not be backed up as the source file(s) no longer exists**.

Key Features

- Identify and remove the files and/or folders in the backup destination(s) which do not appear in the index
- Identify and remove the files and/or folders which appear in the index but do not actually exist in the backup destinations (i.e. AhsayCBS, Cloud storage, or Local storage)
- Identify and remove corrupted files from the backup destination(s) when the **Run Cyclic Redundancy Check (CRC) During Data Integrity Check** setting is enabled
- Identify and remove partially uploaded (orphan) files from the backup destination(s) to free up storage space
- **(TEST MODE) confirmation screen** (applicable on AhsayOBM/AhsayACB client)
- Update storage statistics

Data integrity check CANNOT fix or repair files that are already broken.

Initiating Data Integrity Check (DIC)

Data Integrity Check can be started using the following options:

- AhsayOBM/AhsayACB client GUI
- AhsayCBS Web Console for Run on Server (Office 365 and Cloud File) Backup
- *RunDataIntegrityCheck.bat* batch file (applicable for Windows operating system only)
- *RunDataIntegrityCheck.sh* script file (applicable for FreeBSD/Linux (CLI) operating systems only)

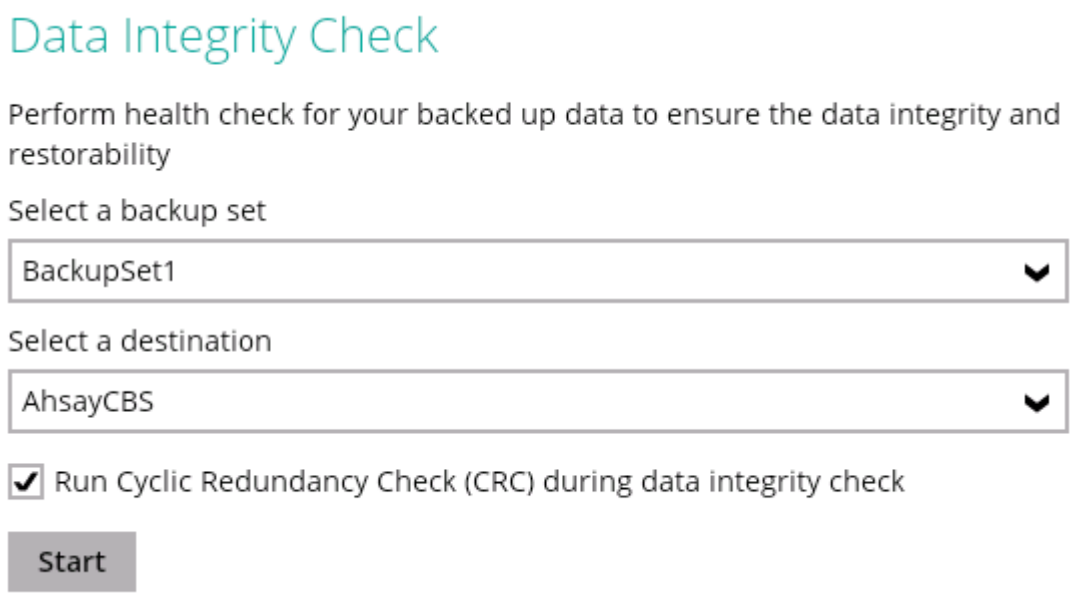
Data Integrity Check (DIC) Modes

There are two (2) data integrity check modes:

- With Run Cyclic Redundancy Check (CRC) disabled (Default mode)
- With Run Cyclic Redundancy Check (CRC) enabled

With Run Cyclic Redundancy Check (CRC) disabled (Default mode)

This is the default setting of the data integrity check. Running a data integrity check on this mode allows the AhsayOBM/AhsayACB client or AhsayCBS backup server to perform a comparison between the files and/or folders on the backup destination(s) and the list of the files and/or folders recorded in the current index file.



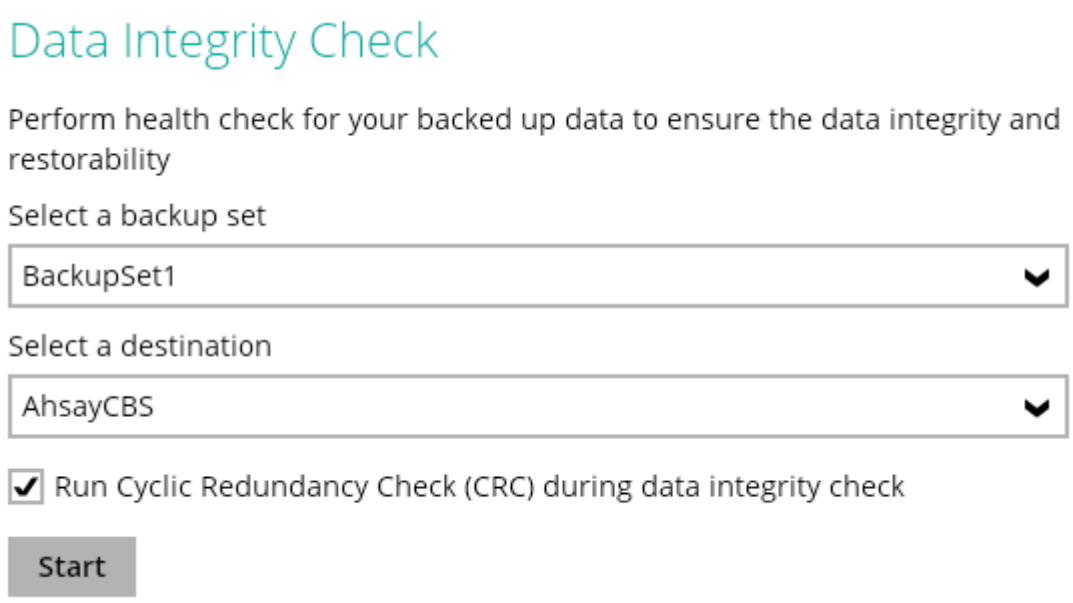
When should I run a Data Integrity Check in default mode?

- If you encounter index issues on your backup/restore job
- If you know or suspect the backup set storage statistics are not updated or incorrect and cannot wait for the next weekly Periodic Data Integrity Check (PDIC) job
- If you need to remove partially uploaded (orphan) files from the backup destination(s) to free up space, as partially uploaded (orphan) files will remain in the backup destination(s) when backup jobs with large files (i.e. database, VMware/Hyper-V, Windows System) backups are terminated unexpectedly or crashes

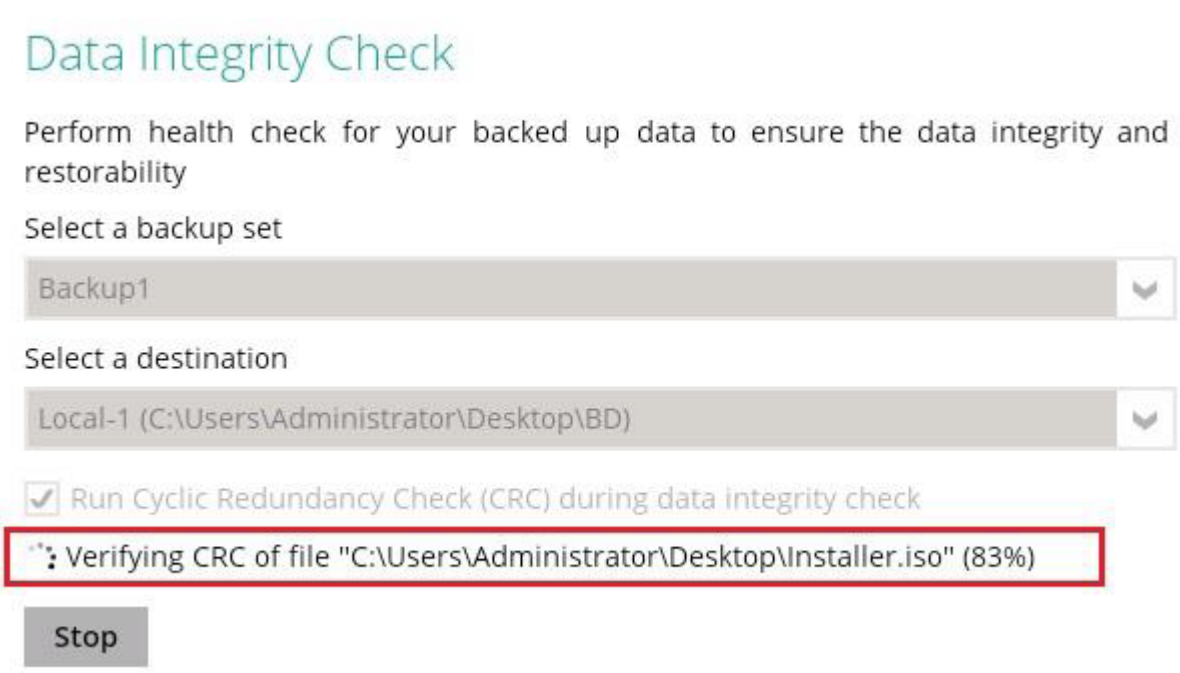
With Run Cyclic Redundancy Check (CRC) enabled

Running a data integrity check on this mode will perform check on the integrity of the files in the backup destination(s) against the checksum file generated at the time of the backup job.

If there is a discrepancy, this indicates that the file(s) on the backup destination(s) are corrupted. The AhsayOBM/AhsayACB client or AhsayCBS backup server will remove these files from the backup destination(s). If these files still exist on the client machine or backup server on the next backup job, The AhsayOBM/AhsayACB client or AhsayCBS backup server will upload the latest copy.



For large file sizes, a **percentage progress** will be displayed throughout the data integrity check job when this setting is enabled:



When should I run a Data Integrity Check with Run Cyclic Redundancy Check (CRC) enabled?

With Periodic Data Integrity Check (PDIC) and post-backup validation features on the AhsayOBM/AhsayACB v8.3.2.11 or above, it is not necessary to frequently run a Data Integrity Check with Run Cyclic Redundancy Check (CRC) enabled. Also, this option can take a long time to complete as the AhsayOBM/AhsayACB will need to download all the files and/or folders from the backup destination(s) on the AhsayOBM/AhsayACB client machine in order to perform the actual Cyclic Redundancy Check (CRC).

To reduce the time taken, you should consider selecting only one backup destination at a time if applicable.

It is recommended to use this option:

- When the AhsayOBM/AhsayACB client machine encounters “corrupted file” errors during a restore job, running a data integrity check with Cyclic Redundancy Check (CRC) enabled may help to identify and clean up the corrupted files and allows the end user to recover any remaining data from the backup set(s).
- When a backup destination has encountered a hardware failure (e.g. a disk failure on an AhsayCBS user home drive or AhsayOBM/AhsayACB Local destination drive).

If the AhsayOBM/AhsayACB client machine is accessing the internet on a metered internet connection plan, it will incur additional data charges from your ISP (Internet

Service Provider) as a result of the data download.

If the backup destination(s) are commercial Cloud Storage destinations, it may incur additional charges from your Cloud Storage Provider as a result of the data download.

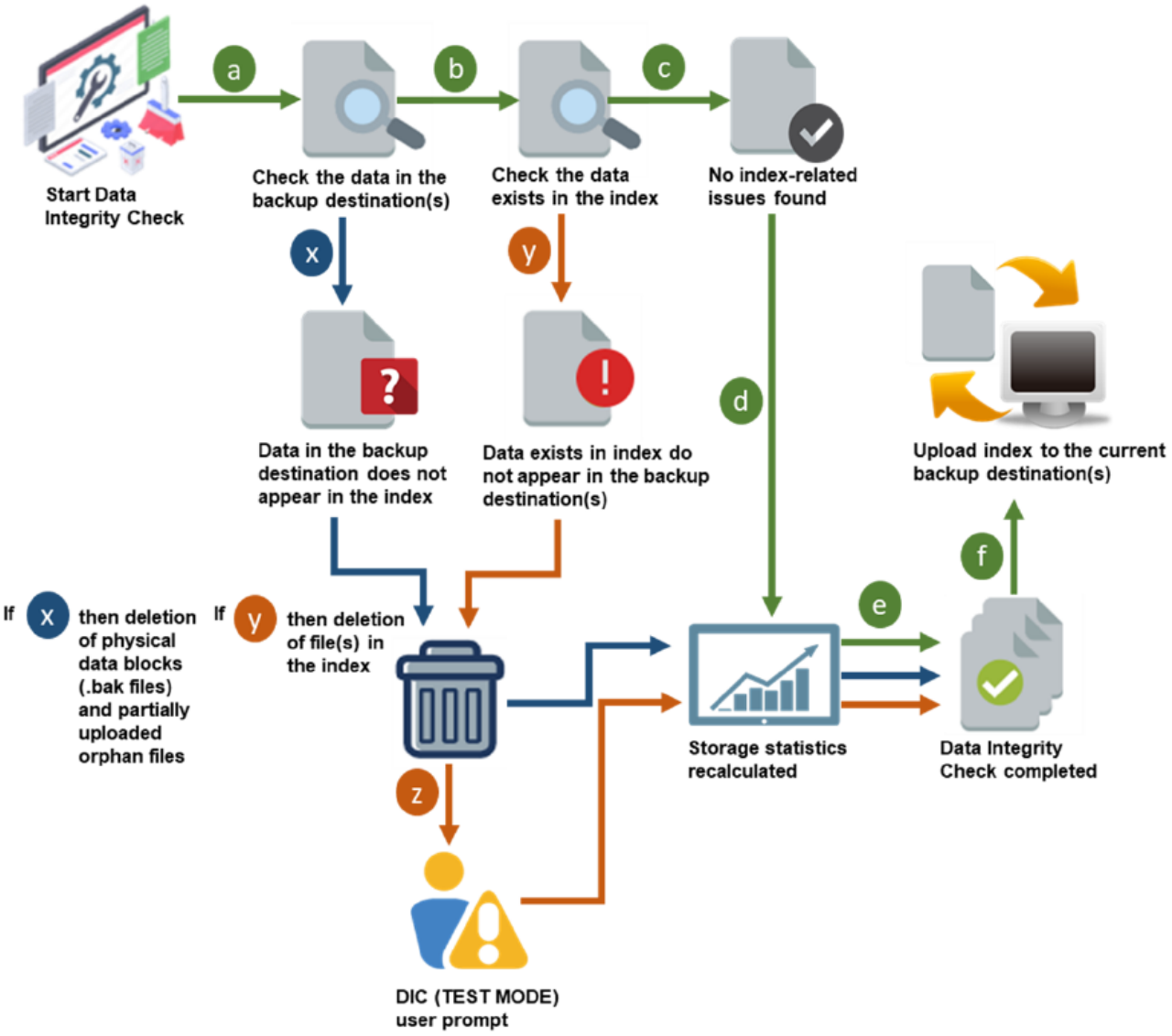
Limitations

- Data Integrity Check has to be started manually from the AhsayOBM/AhsayACB client UI. It cannot be remotely started from the AhsayCBS web console or scheduled backup to run automatically. The only exception is for a Run on Server (Office 365 or Cloud File) backup sets were a data integrity check can be started from the AhsayCBS web console
- When a Data Integrity Check has identified issues on the backup set, it may require the end user to confirm the changes before it takes the corrective actions
- When a data integrity check is running, a backup and restore job cannot be run and vice versa: When an active backup or restore job(s) is running, a data integrity check cannot be run

How It Works

The following diagrams show the detailed flow for each data integrity check mode.

With Run Cyclic Redundancy Check (CRC) disabled (Default mode)



a Check the data in the backup destination(s) if they appear in the index

- → If **YES**, proceed to b
- → If **NO**, proceed to x

b Check the data in the index if they actually appear in the backup destination(s)

- → If **YES**, proceed to c
- → If **NO**, proceed to y

c No index-related issues have been found

d Storage statistics of data area and retention area are recalculated

e Data Integrity Check is complete

f Indexes will be uploaded to the current backup destination(s)

x Physical data blocks (.bak files) that do not exist in the index and partially uploaded orphan files will be automatically removed from the backup destination(s)

Proceed to d

y File(s) in the index that do not actually exists in the backup destination(s) will be deleted

- → If either of the **criteria's** matches, proceed to z
- → If **NOT**, proceed to c

z (TEST MODE) confirmation screen will prompt user to proceed with the corrective actions (recommended)

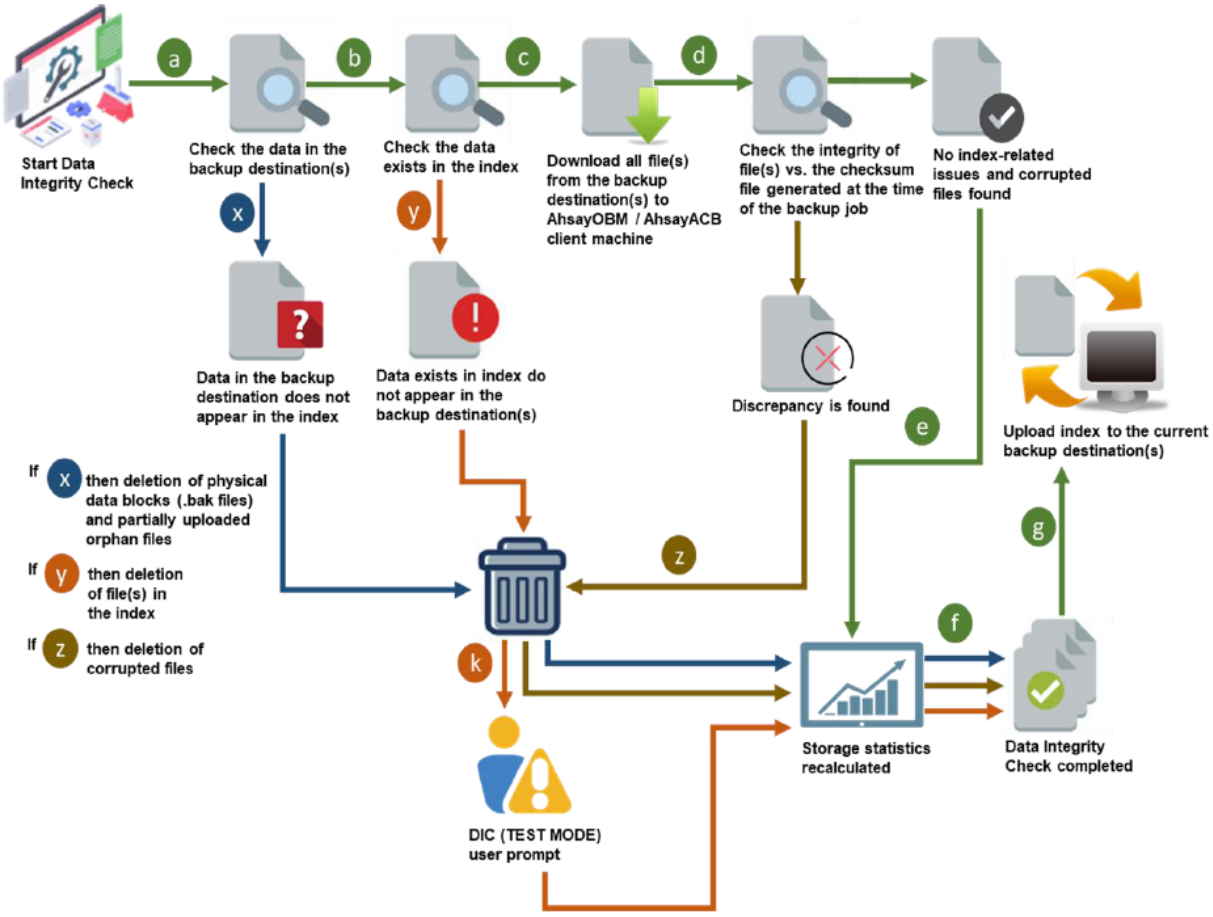
- → If the user selects **YES**, then changes will be applied
- → If the user selects **NO**, then the data deletion will be discarded

Proceed to d

By default, (TEST MODE) confirmation screen will only prompt if either of the **criteria's** below matches the backup data:

- deleted number of backup files is over 1000
- deleted number of backup file size is over 512 MB (in total)
- deleted number of backup files is over 10% of total backup files

With Run Cyclic Redundancy Check (CRC) enabled



a Check the data in the backup destination(s) if they appear in the index

→ If **YES**, proceed to b

→ If **NO**, proceed to x

b Check the data in the index if they actually appear in the backup destination(s)

→ If **YES**, proceed to c

→ If **NO**, proceed to y

c For Run on Client (agent-based) backup sets, files in the backup set are downloaded from the backup destination(s) to the AhsayOBM/AhsayACB client

For Run on Server (agentless) backup sets, proceed to d

d Check the integrity of the files in the backup destination(s) against the checksum file generated at the time of the backup job

→ If any discrepancy is **FOUND**, proceed to z

→ If **NONE**, proceed to e

e Storage statistics of data area and retention area are recalculated

f Data Integrity Check is complete

g Indexes will be uploaded to the current backup destination(s)

x Physical data blocks (.bak files) that do not exist in the index and partially uploaded orphan files will be automatically removed from the backup destination(s)

Proceed to e

y File(s) in the index that do not actually exists in the backup destination(s) will be deleted

→ If either of the criteria's matches, proceed to k

→ If **NOT**, proceed to d

z Corrupted files will be automatically removed from the backup destination(s)

Proceed to e

k (TEST MODE) confirmation screen will prompt user to proceed with the corrective actions (recommended)

→ If the user selects **YES**, then changes will be applied

→ If the user selects **NO**, then the data deletion will be discarded

Proceed to e

By default, (TEST MODE) confirmation screen will only prompt if either of the criteria's below matches the backup data:

- deleted number of backup files is over 1000
- deleted number of backup file size is over 512 MB (in total)
- deleted number of backup files is over 10% of total backup files

Test Mode Confirmation Screen

Normally aspart of the data integrityjob, **(TEST MODE)** confirmation screen is usually displayed once a data integrity check is completed, which gives a summary report of the corrupted files, invalid indexes,or storage statistics issue for each backup destination. The (TEST MODE) confirmation screen allows the enduser to review the results of the data integrity check, and to decide whether they would like to proceed with the corrective actions.To further streamline the data integrity check process and improve user experience,the (TEST MODE) confirmation screen will **ONLY** prompt if either of the criteria's below matches the backup data during the data integrity check operation:

- deletednumber of backupfiles isover 1,000
- deletednumber of backup file size isover 512MB (in total)
- deletednumber of backup files isover 10% ofthetotal backup files

Otherwise, the data integrity check job will **automatically** take corrective actions.

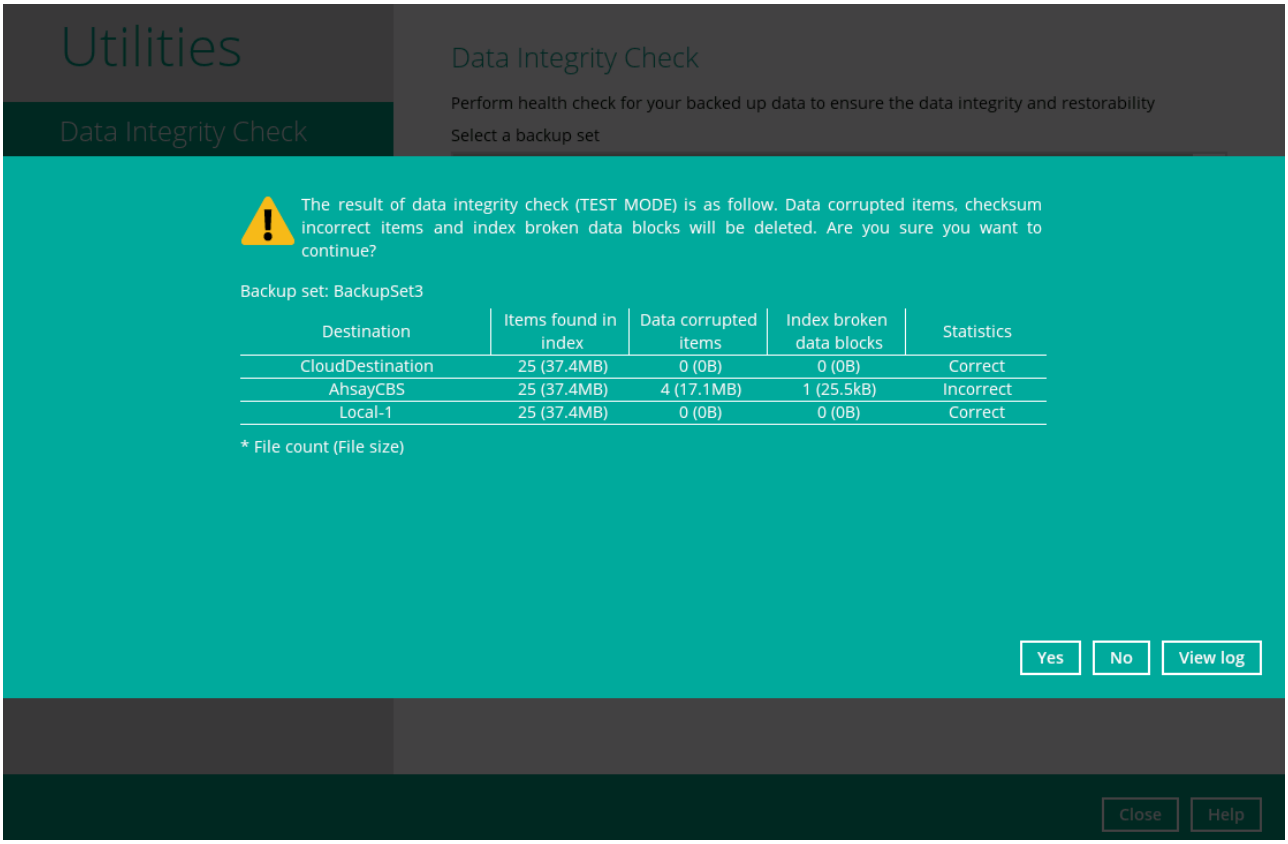
The (TEST MODE) screen includes five (5) summary report for the following items found per backup destination:

Items	Description
Destination	this indicates the destination of the backup data where the data integrity check will be run
Items found in index	the number of files and its total size (MB) that appear in the index
Data corrupted items	the number of files and its total size (MB) found to be corrupted
Index broken data blocks	index with its associated data blocks which found to be corrupted
Statistics	storage statistics status of the data area and retention area of the backup destination (i.e. correct or incorrect status)

> Although you select ALL backup sets before starting the data integrity check, the (TEST MODE) confirmation screen will prompt one at a time with the corresponding backup set(s).
>
> For example, the data integrity check has run with three (3) backup sets and all these backup sets match the criteria's of the (TEST MODE) confirmation screen, the (TEST MODE) confirmation screen will prompt three times to confirm if the end user will take corrective actions for the three backup sets.

Below is an example of a (TEST MODE) confirmation screen with the following scenario:

- Multiple backup destinations, corrupted items and index-related issues found with correct and incorrect storage statistics.



How does Data Integrity Check (DIC) compare with Periodic Data Integrity Check (PDIC)

Periodic data integrity check is performed at the beginning of a backup job, which provides an additional regular data integrity check of the backup data and updates the storage statistics for each backup set. The PDIC feature is enabled from v8.3.2.11 or above and cannot be turned off. This is to ensure a maximum protection of the backup data.

Unlike with the Data Integrity Check (DIC), the PDIC starts automatically and performs a quick check of all the backup destination(s) without the end user intervention.

The PDIC will be initiated automatically once either of the following conditions is met:

- Will be triggered on a weekly basis, usually on the first run of backup job that falls on any of these days: Friday, Saturday, or Sunday
- *OR**
- If there is no active backup job(s) running on Friday, Saturday, or Sunday, then the PDIC will be triggered on the next available backup job

E.g. If the last PDIC job was run more than seven (7) days ago, then the subsequent PDIC job(s) will run seven days from that day onwards.

Features	**Data Integrity Check (DIC)**	**Periodic Data Integrity Check (PDIC)**

Features	Data Integrity Check (DIC)	Periodic Data Integrity Check (PDIC)
Runs automatically	X	✓
Allows selection of backup destination(s)	✓	X
Run Cyclic Redundancy Check (CRC) feature	✓	X
Identify and remove the files and/or folders in the backup destination(s) which do not appear in the index	✓	✓
Identify and remove the files and/or folders which appear in the index but do not exist in the backup destination(s)	✓	X
Identify and remove partially uploaded (orphan) files from the backup destination(s)	✓	X
(TEST MODE) confirmation screen feature	✓	X
Update Storage Statistics	✓	✓

FAQs

- How to run a Data Integrity Check on Linux (CLI) machine

Issues

- Index file for destination of backup set is found to be corrupted (User prompted to delete all data when performing data integrity check)

From:
https://wiki.ahsay.com/ - **Ahsay Wiki**

Permanent link:
https://wiki.ahsay.com/doku.php?id=public:data_integrity_check&rev=1588044740

Last update: **2020/04/28 11:32**