FAQ: How to run a Data Integrity Check for backup data stored in backup destination?

Article ID: 8004
Reviewed: 20/03/2020

Product Version:
AhsayOBM / AhsayACB: 8.1 or above
OS: Windows, MacOS, Linux(GUI), QNAP, Synology

Description

This article outlines the steps to perform a Data Integrity Check for backup data stored in the backup destination (e.g. local destination, AhsayCBS or other cloud storage).

The functions of the Data Integrity Check (DIC) is to:

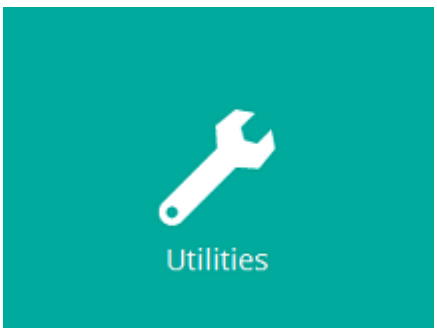
1. Identify and remove the files and/or folders in the backup destination(s) which do not appear in the index.
2. Identify and remove the files and/or folders which appear in the index but do not actually exist in the backup destination(s).
3. Identify and remove corrupted files from the backup destination(s) when the Run Cyclic Redundancy Check (CRC) During Data Integrity Check setting is enabled.
4. Identify and remove partially uploaded (orphan) files from the backup destination(s) to free up storage space.
5. Update the storage statistics for the backup set(s).

The Data Integrity Check CANNOT fix or repair files that are already corrupted. It will only identify and remove any corrupted files from the backup destination(s), so these files can be re-uploaded again on the next backup job if they still exist on the backup source. However, any corrupted files removed from the Retention Area will be not be re-uploaded and will no longer be restorable.

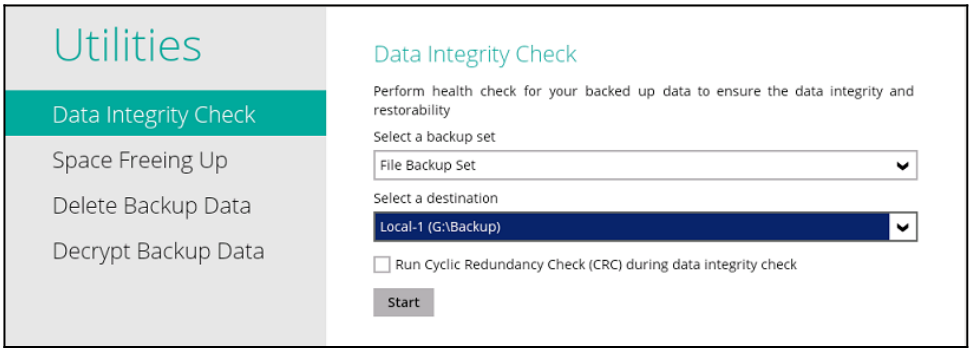
Steps

To perform a data integrity check, login to the AhsayOBM / ACB user interface:

1. Login to the AhsayOBM / ACB user interface, select the **Utilities** tile.



2. Select the corresponding **backup set** from the dropdown menu and then select the **backup destination** from the dropdown menu



Note:
The "All" backup sets option will enable check for all destinations of all backup sets. The time required to complete the integrity check will depend on the number of backup sets and destinations.

3. Enable the **Run Cyclic Redundancy Check (CRC) during data integrity check** option to check on integrity of the files against the checksum file generated at the time of the backup job.

4. Then press the **Start** button to start the data integrity check.

Note:
i. A data integrity check can only be performed when there is no manual / scheduled backup job in progress (of the corresponding backup set). It is highly recommended to temporarily disable the backup schedule to ensure that no scheduled backup is started while the data integrity check is still running.

The following error message will be displayed to indicate that the data integrity check had skipped a backup set with active backup job

Skipped Backup Set="Backup Set". Reason = "Backup Job "Backup Set" is still running."
Finished data integrity check with error on backup set "Backup Set (Backup Set ID)

ii. The time required to complete a Data Integrity Check depends on a number of factors, such as the number of files / folders in the backup set(s), bandwidth available on the client computer, hardware specifications of the client computer such as the disk I/O and CPU performance, and if there are other resource intensive job running.

iii. If the CRC (Cycle Redundancy Check) option is enabled, backup data will be streamed from the backup destination (e.g. the cloud storage location or FTP location for example), to the client computer in order to perform the CRC check. This may incur additional charges from your Cloud Storage provider. For users with metered internet connection, this could result in additional charges by your ISP (Internet Service Provider).

iv. During a data integrity check, pay attention to the resource usage on the client computer.

v. For backup destinations on cloud storage service, such as Amazon S3, data integrity checks are expected to be performed by the cloud storage service provider.

According to Amazon S3 FAQs (http://aws.amazon.com/s3/faqs/)

> Q: How is Amazon S3 designed to achieve 99.999999999% durability?
>
> Amazon S3 redundantly stores your objects on multiple devices across multiple facilities in an Amazon S3 Region. The service is designed to sustain concurrent device failures by quickly detecting and repairing any lost redundancy. When processing a request to store data, the service will redundantly store your object across multiple facilities before returning SUCCESS. Amazon S3 also regularly verifies the integrity of your data using checksums.
>
> Q: What checksums does Amazon S3 employ to detect data corruption?
>
> Amazon S3 uses a combination of Content-MD5 checksums and cyclic redundancy checks (CRCs) to detect data corruption. Amazon S3 performs these checksums on data at rest and repairs any corruption using redundant data. In addition, the service calculates checksums on all network traffic to detect corruption of data packets when storing or retrieving data.

Consult with your cloud service provider to ensure that CRC checks are performed regularly for your data.

Keywords

dic, index, rebuilding, bdb, backupset, set, perform