

Resilient Change Tracking

Shwetank Singh Yadav¹, Swarupa Gudsoorkar², Umang Kelani³, Vikas Kashyab Jha⁴,

Dr. Mrs. N. A. Deshpande⁵, Miss Suvarna Kadam⁶

^{1,2,3,4}Student, Computer Department of D. Y. Patil College of Engineering, Akurdi, Maharashtra, India, 44.

⁵Head of the Computer Department of D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India, 44.

⁶Assistance Professor, Computer Department of D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India, 44

Abstract - This paper presents a survey on the applications of data backup on the virtual platform and implementation library based on Resilient Change Tracking. We put forth the discussion on combined use of state-of-the-art Resilient Change tracking (RCT), Hyper-V and virtualization, which is capable of providing seamless integration with the backup application for the functionalities of incremental backup. We present brief report on the novel methods used in backup of data

Key-Words: Backup, Incremental Backup, Windows Server 2016, Hyper - v 2016, Checkpoints, RCT

1. INTRODUCTION

In today's world data, time and resource are everything. Backing up data has become significant due to its high importance in various fields. Various events like system crash, accidental deletion, or data corruption constantly impose threats to the consistency of data. So data backup is necessary. That being said, consider that you just did a first full backup of your computer. Some time passed, and you want to back it up again. How much data changed since then? Probably not a lot. Does it make sense to back up everything again? Not really. Why would you want to spend time and resources on copying something you already backed up? RCT is the solution to this problem.

1.1 Motivation of the Project

In an organization, information stored on your computer might be enormously momentous. Hence, regularly scheduled Backup are required, as sequential Backup of the data are being done in order to provide speedy recovery during a disaster. Therefore, every time it is not feasible to take a full backup. Once the first full backup of Virtual Machines is done, the consequent incremental backup must transfer only the changed blocks of data rather than sending whole VM data again. Change Block Tracking (CBT) technology has helped to address this issue. On a physical server, disk-imaging backup solutions install a small driver to track changes to the blocks on the disk. As the Microsoft Hyper-V server do not have CBT as an inbuilt feature, with the release of Windows Server 2016, Microsoft introduced the Resilient Change Tracking (RCT) technology to Hyper-V.

1.2 Terminologies

1.2.1 Backup:

Backup is the process of copying physical or virtual files to a secondary site for protection in case of equipment failure or other catastrophes. Types of Backup are as follows.

Full backup: The files and folders everything that needs to be backed up will be backed up in their entirety. If the full backup method is chosen as a default, the next time you backup your data, the files and folders will be backed up again entirely.

Advantages of Full Backup:

- Fast easy recovery as complete data is readily available.

- Files folders are backed up to one backup set.

Disadvantages of Full Backup:

- More storage space is needed.

- Additional bandwidth required.

- Time-consuming if full backup all the time.

Incremental Backup: Backs up all of the files that have changed since the last backup, whether the backup was full or incremental. For instance, if a full backup was made on Saturday night, then backup may be performed on Tuesday night to backup files that have changed since Saturday night. On Wednesday night, another incremental backup is performed to backup files that have changed since Monday night. The motive of incremental backup is to reduce the time interval between Backup, requiring fewer data to be backed up.

Advantages of Incremental Backup:

- Fast backup windows, as there is less data as compared to full backup.

- Less storage space needed.

- Allow retention of several versions of same files.

Disadvantages of Incremental Backup:

- Slower recovery, as all increments, must be restored.

- Initial full backup is needed before incremental Backup start.

- A full backup and all incremental Backup are needed for recovery.

Differential Backup: Backs up files that have been changed since the last full backup. For instance, if a full backup was performed on Saturday night, then on Monday night, differential backup will back up all of the files that have changed since Saturday night. On Wednesday night, differential backup will Monday and Wednesday and so on.

Advantages of Differential Backup:

- Less storage space needed as compared to incremental.

- Only full backup and the last differential backup needed for Backup.

- Less storage space needed.

- Allow retention of several versions of same files.

Disadvantages of Differential Backup:

- Slower Backup than incremental Backup.

- Initial full backup is needed before differential Backup start.

- A full backup all differential Backup are for recovery

Common backup scenarios	Media Space Required for one Month (20 TB @ 5% daily rate of change)	Media required for recovery
Full daily (weekdays)	Space for 22 daily full (22 * 20 TB) = 440.00 TB	Most recent backup only
Full (weekly) + Differential (weekdays)	Full, plus most recent differential since full (5 * 20 TB) + (22 * 5% * 20 TB) = 124.23 TB	Most recent full + most recent differential
Full (weekly) + Incremental (weekdays)	Full, plus all incremental since weekly full (5 * 20 TB) + (22 * 5% * 20 TB) = 122.00 TB	Most recent full + all incremental since full

1.2.2 Virtualization:

Virtualization is the idea of creating a virtual rather than real or genuine version of some-thing, such as an operating system, a server, a storage device or network resources.

Hypervisor: A Hypervisor, also known as Virtual Machine Monitor (VMM) is an-other technology at the core of system virtualization. A hypervisor provides the foundations for virtualization management, which includes policy-based automation, virtual hard disk, lifecycle management, live migration and real-time resource allocation. At the center of virtualization is a hypervisor, a layer of software that manages the creation and execution of virtual machines. Hyper-V is the type-1 hypervisor for windows platform and it is used for virtualization. Types of hypervisor are,

Type 1 hypervisors: Type 1 hypervisors are those that run directly on the system hardware and offers a higher level of virtualization efficiency and security. Oracle OVM, ESXi, Hyper-V, KVM.

Type 2 hypervisors: Type 2 hypervisors are those that run on a host operating system that provides services, such as I/O device support and memory management. Example VMware fusion , oracle virtual box, oracle VM for x86

2. RELATED WORK

Earlier during the time of full backup, all data blocks are copied to the repository. For the subsequent backup we only want the changed block to be backed up again, but in case of no changes involved, the source VM needs to be scanned and hash value is calculated for each data block and then the two VM's(Old and the new one) are compared based upon their hash value. Although it provides a great saving on the storage as we do not need to copy everything each time a scanning is done. One major drawback of this technique is that it requires entire VM scan i.e. Rescanning each block hash thus utilizing both resources and time.

2.1 Delta Differencing

Delta Differencing is a technique, which uses incremental backup perspective, and only the changed blocks after the first full backup are taken as a backup. The changes are stored in files called deltas.

The process requires scanning a backup file set and marking the blocks that have changed since the last backup period. Changed block, rather than the original file set, are sent to the backup target. Due to the backup of only those data blocks, which contain the differences, it is possible to perform rapid and more frequent backup cycles without taking the full control of the bandwidth. Consider a scenario where an initial backup or disaster recovery set of 30 TB takes a full day, for example, a typical delta difference of 10 GB per day can be transferred in just a few hours. Delta differencing, which is sometimes referred to as a delta differential backup, is frequently used in WAN-based backups, cloud backups and virtual machine snapshots

2.2 VMware Snapshot

VMware Snapshot are a point in time replica of the virtual machine's disk file (VMDK). Snapshots contains a change log for the virtual disk, which are used to restore to a particular point in time when a system failure occurs. VM snapshot typically represent the state of any virtual machine. A regular practice of snapshot backup process can notably reduce the recovery point objective for the protective VM. For example, consider that you take VM snapshot every 20min; there is loss of upto 15min of data in the event of system failure. Thus, snapshot backup is an important tool for data-protection for VM.

2.3 CBT

One of the ways of efficiently implementing it was in the form of Change block tracking (CBT). It works on bitmaps i.e. it creates bitmap of all the blocks on a virtual hard drive. As soon as it observes any data changes in a particular block, the respective block is marked as "changed". At the time of backup, it tracks the bitmap of the entire changed block since the last backup and only those blocks are backed up.

Drawbacks:-

No inbuilt or native feature available in windows so backup vendors to monitor changes used "filters drivers".

All blocks that are being monitored are stored in memory bitmap on the host. Therefore, if anything happens to memory such as VM migration or power crash, then the whole bitmap is lost so we will not be able to retrieve those changes.

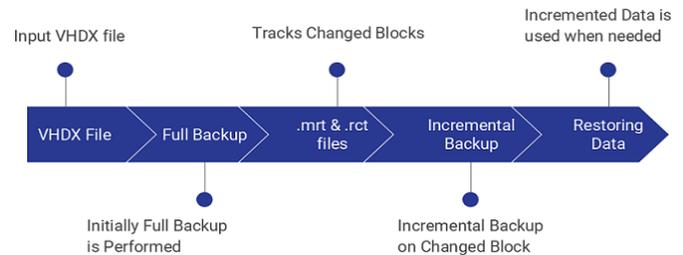
3. PROPOSED WORK

3.1 Resilient Change Tracking (RCT):

Microsoft solves the earlier potentials issues which were present in CBT with a new incremental backup technique i.e. RCT which is a built in feature in Windows server 2016. RCT creates three Bitmaps: one in memory and two on disk. The bitmap one in the memory is same as that of CBT and it will track the bitmap and copy only the changed block when an incremental backup is done. The remaining two bitmaps i.e. (.MRT and .RCT file) are used during the time of VM migration or when there is a sudden power loss. The RCT file will be used in normal operation i.e. during bitmap Loss when VM is migrated from one host to another and MRT file will be used in case of power outage to recover bitmaps of changed block. It offers easier, faster incremental backup and there is no need of third party filters.

Checkpoints: One of the great benefits of virtualization is the ability to easily save the state of a virtual machine. In Hyper-V, this is done with virtual machine checkpoints. Virtual machine checkpoints (formerly known as virtual machine snapshots) capture the state, data, and hardware

configuration of a running virtual machine. Checkpoints provide a fast and easy way to revert the virtual machine to a previous state. A library which works on RCT and seamlessly integrates with any backup application and provides the changed blocks since the last backup so as to only backup those blocks and automate the whole process.



Advantages:

- Works with any windows backup application.
- Easy backup & restore.
- 3 bitmaps (1 memory, 2 disks).
- No 3rd party filter needed.

4. WORKFLOW

Step 1: Attach Virtual Disk Initial step include automatic attachment of virtual Disk for performing data backup operations.

Step 2: Virtual Disk Info VHDX file is taken as input and complete information is retrieved using GET_VIRTUAL_DISK_INFO().

Step 3: Checkpoint Creation VHDx Checkpoint creation using win32 API in PowerShell script and also one console application which will call that script.

Step 4: Dynamic Library Creation of .DLL file that will integrate with any

Backup application and helps in Backup and Restore.

Step 5: RCT ID Retrieval of RCT id and using this id the changed blocks are identified and backed up accordingly.

Step 6: Restore & Backup Finally application will provide the functionality of full and incremental backup and afterwards restoration of data can be done when needed.

5. CONCLUSION

In today's world data, time and resource are everything. Backing up data has become significant due to its high importance in various fields. Various events like system crash, accidental deletion, or data corruption constantly impose threats to the consistency of data. So data backup is necessary. For this purpose, RCT library can be used, which takes the incremental backup of the data.

REFERENCES

- [1] Nakamura, S., K. Nakayama, and T. Nakagawa. "Optimal backup interval of database by incremental backup method." *Industrial Engineering and Engineering Management*, 2009. IEEM 2009. IEEE International Conference on. IEEE, 2009.
- [2] Ha, Viet Hai, and Eric Renault. "Discontinuous incremental: A new approach towards extremely lightweight checkpoints." *Computer Networks and Distributed Systems (CNDS), 2011 International Symposium on*. IEEE, 2011.
- [3] <https://docs.microsoft.com/enus/virtualization/hyper-v-on-windows/userguide/checkpoints>
- [4] [https://technet.microsoft.com/enus/library/dn818483\(v=ws.11\).aspx](https://technet.microsoft.com/enus/library/dn818483(v=ws.11).aspx)
- [5] <http://www.enterprisefeatures.com/backup-types/full-incremental-differential/>
- [6] <https://www.vembu.com/blog/resilient-changetracking-microsoft-hyper-v/>
- [7] <https://www.acronis.com/enus/blog/posts/leveraging-hyper-v-2016-resilient-changetracking-backup>