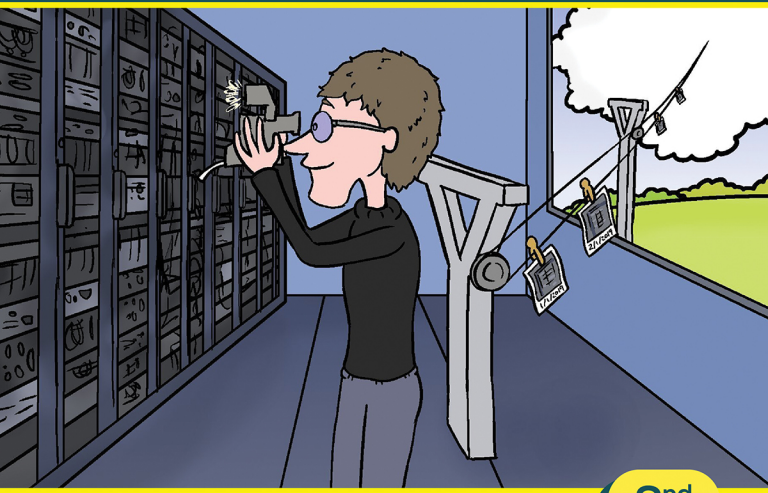
ConversationalGeek

Conversational
Azure Data Protection

By Wayne Dipchan (MCSE Cloud Platform and Infrastructure)



Learn about:

- Native Azure backup options
- Azure Backup pros and cons
- Areas where third-party backup solutions make sense

2nd MINI Edition

Sponsored by

COHESITY

Sponsored by Cohesity

Cohesity radically simplifies data management. We make it easy to protect, manage, and derive value from data – across the data center, edge and cloud. We offer a full suite of services consolidated on one multicloud data platform: backup and recovery, disaster recovery, file and object services, dev/test, and data compliance, security, and analytics – reducing complexity and eliminating **mass data fragmentation**. Cohesity can be delivered as a service, self-managed, or provided by a Cohesity-powered partner.

COHESITY

For more information, visit www.cohesity.com

Conversational Azure Data Protection (2nd Mini Edition)

by Wayne Dipchan

ConversationalGeek®

Conversational Azure Data Protection (2nd Mini Edition)

Published by Conversational Geek® Inc.

www.ConversationalGeek.com

Trademarks

Warning and Disclaimer

Additional Information

For general information on our other products and services, or how to create a custom Conversational Geek book for your business or organization, please visit our website at www.ConversationalGeek.com.

Publisher Acknowledgments

All of the folks responsible for the creation of this book:

Author:	Wayne Dipchan
Project/Copy Editor:	Pete Roythorne
Content Reviewer(s):	Douglas Ko

The “Conversational” Method

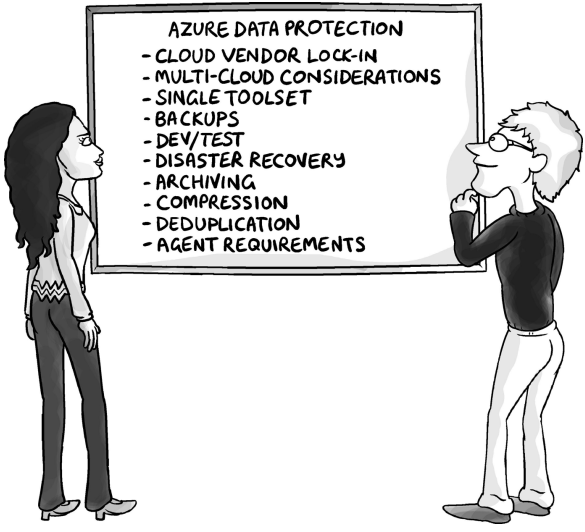
We have two objectives when we create a “Conversational” book. First, to make sure it’s written in a conversational tone so that it’s fun and easy to read. Second, to make sure you, the reader, can immediately take what you read and include it into your own conversations (personal or business-focused) with confidence.

“Geek in the Mirror” Boxes

We infuse humor and insight into our books through both cartoons and light banter from the author. When you see one of these boxes it’s the author stepping outside the dialog to speak directly to you. It might be an anecdote; it might be a personal experience.

Within these boxes I can share just about anything on the subject at hand from the author or other SMEs. Read ’em!

Azure Data Protection



Data Protection in its simplest terms is the process of duplicating data to another location… could we pick a more boring topic?

Really, if we think about it, no one wants to back up data. An organization only backs up data because they feel they must. Backing up data is like an insurance policy; that is paid for just because it may

be needed one day in the future. In many cases, a loss of data or the inability to access data for an extended period of time could have a catastrophic impact on a business. This makes data valuable. And it is the reason why some not-so-nice individuals try to hijack data by means of ransomware, holding data hostage until they get paid. In addition to the value of the data, there may be regulatory requirements that need to be met regarding the retention of data backups.

The Evolution of Backup

Believe it or not, there was a time when all servers where physical. You may remember a tape drive installed directly on the server and backup software running every night to copy your data onto a tape. Someone had to change the tape every day and make sure to take it offsite. In larger environments, you may have had a central tape library.

You may still have a tape library with physical tapes. If so, it is time to start thinking about upgrading your solution.

Some vendors offer technologies like Virtual Tape Libraries (VTL) and the ability to “Air Gap” data. Server virtualization has enabled businesses to make copies of entire servers with ease. As cloud adoption progresses, backup to the cloud is becoming more and more attractive as a backup solution due to its “as a service” approach.

Air Gapping places your data on an isolated network once it is copied. Offline Tape backups are also a form of air gapping albeit a solution that is becoming less popular. Air Gapping protects the copied data from any corruption that may affect production data, as it cannot be replicated to the air-gapped data.

Here are some items that should be considered with a backup strategy:

1. Defined DR Tier levels for Applications and Data per criticality

    a. Each Tier designation should have a specific recovery point objective (RPO) and recovery time objective (RTO). (See note below)
    b. Every application should be assigned a DR Tier.
2. Software that can orchestrate server replication and data backup.
    a. Monitor replication/backup
    b. Test recovery of data
    c. Deduplicate and compress data
    d. Granular restoration (single file if need be)
    e. Make backup data available for use
    f. Move archive data to lower cost storage
    g. Failover workloads to DR site and failback to primary site
    h. Group backups to ensure application consistent restores
    i. Manage IP addressing updates if needed
    j. Encryption of data in transit and at rest
    k. Making backed up data immutable

l. Similar workflow for on-premises and in the cloud

Note: Recovery time objective refers to the amount of time allowed for the data to become available after a disaster has occurred. Recovery point objective refers, in time increments, to how much data can be lost. If the last 15 minutes is the only amount of data loss acceptable, then a copy of that data would need to be made every 15 minutes or less. Can all application backups be configured at the lowest possible RPO? Yes, this outcome is achievable with "Continuous Data Protection" (CDP); however, you should always have a clear understanding of what RPO is required and configure your backup and monitoring accordingly.

Native Azure Backup Options

Azure includes a built-in "free" option to back up virtual machines (VMs) through the Azure Portal. This backup is configured manually with an editable policy that will only allow you to select either a daily or weekly frequency and a few options on backup retention (how long you want to keep the backup copy). The best RPO this solution offers is 24 hours.

It is considered “free” because it is technically an included feature when an Azure VM is deployed, however, you are still going to pay for the storage used. This is a solution that does not compress or deduplicate, so there ultimately is a cost. It is important to note that this is only for Azure VMs. Microsoft knows this is not an enterprise-grade solution, so they offer another paid-for option called Azure Backup.

Azure Backup

Because Microsoft knows the VM built-in solution is not a prime-time enterprise approach to data protection, Azure Backup is a solution with which you pay for both the service and the storage costs.

When paying for a solution (provided by Microsoft or a third party), it’s important to determine which solution is the most cost effective and efficient for the organizations’ needs. Compare cost to feature-set before deciding.

Azure Backup offers

1. A fairly basic solution to move or initiate backups to the Azure cloud
2. Backup Azure Infrastructure as a Service (IaaS) VMs
3. Leverage the scalability of the Azure cloud for backups by adding additional agents on the VMs
4. Unlimited data transfer—inbound and outbound. This is to and from an Azure Recovery Services Vault and more specifically, outbound refers to data transferred during a restore operation.
5. Secure Data
6. Application consistent backups
7. Short-term and long-term data retention
8. Automatic storage management – Azure backup data resides in the Azure Recovery Services Vault, which is managed by Microsoft and automatically replicated using LRS, ZRS, or GRS (defined below). Data can be moved to on-premises storage (as a primary) which would reduce the cloud expense, but there would be a charge

to move the data out of Azure (unless the move occurs during a recovery action).

Azure storage will be automatically provisioned and charged on a pay as you go basis. An organization will only pay for the storage used. Backup data will also be replicated. There are three options: Locally Redundant Storage (LRS), Zone redundant storage (ZRS), or Geo-redundant Storage (GRS). LRS will make three copies of data in a storage scale unit within a datacenter. ZRS replicates your backup data into availability zones, separate datacenters within a region. In these two scenarios, all copies of data reside in the same region. GRS will replicate data to another region, (and actually make the data read accessible GRS-RA) which will be at least 100 miles away from the primary data center. GRS is the default option; it will protect data from a regional outage. GRS will cost more than LRS or ZRS.

Note: A storage scale unit is a collection of racks with storage nodes. Within this collection of racks, there are Fault Domains (FD) and Upgrade Domains (UD). A fault domain can be thought of as all the storage nodes within a single rack. An upgrade domain is a group of storage nodes that will be

upgraded together. The three copies of data in LRS are spread across FDs and UDs so the data is resilient to a fault on a single rack or an update that is being applied.

In addition to Azure Backup, there is another solution Microsoft offers to assist with data protection when disaster strikes. This is Azure's disaster recovery as a service (DRaaS) called Azure Site Recovery.

Azure Backup vs Azure Site Recovery

Azure Backup and Azure Site Recovery (ASR) are two separate solutions, they do, however, complement each other to achieve overall Business Continuity and Disaster Recovery (BCDR) goals.

Azure Backup has the ability to back up both on-premises machines and Azure resources. The options available in Azure at the time of this writing include Virtual Machines, Azure File Shares, SQL Server in Azure VM, and SAP HANA in Azure VM. On premise options include Files and Folders, Hyper-V VMs, VMware VMs, SQL Server, SharePoint, Exchange, System State, and Bare Metal Recovery.

We should note that there is also an option to backup resources running on Azure Stack.

Azure Site Recovery is a disaster recovery solution for on-premises machines (physical or virtual) and Azure VMs. This service performs a continuous replication of your servers from one data center to another and orchestrates the failover and failback of your servers.

Continuous replication allows for low RPO, because the delta between the source and replica copy is very small.

Our focus in this book is Azure data protection options. As was mentioned before, Azure Backup can back up both on-premises resources and Azure resources. However, the methodology used in each scenario differs depending on what you are trying to accomplish… let me explain.

To backup on-premises VMs, Hyper-V and VMware, requires the use of the Microsoft Azure Recovery

Services (MARS) agent and/or a backup server. Typically, this would be either a Systems Center Data Protection Manager (DPM) server or a Microsoft Azure Backup Server (MABS). The on-premises machines will back up to this server, and the server would send the backup data to Azure Backup Recovery Services Vault.

You should note that when using the MARS agent, whether on-premises or in Azure, Linux machines are not supported. You must use a backup server, DPM, or MABS to back up Linux guest machines running on either Hyper-V or VMware. Linux host machines are not supported. Remember all VMs require an agent.

Azure Backup can back up both virtual and physical on-premises machines. An installation of the MARS agent will be needed to backup files, folders, and system state on each machine.

When backing up Azure IaaS VMs, you have three options:

1. Enable backups on each individual VM. Azure will install an extension on the VM agent. The agent will then back up the entire VM.

2. For a more granular backup, install the MARS agent on each Azure VM.

3. Use a DPM or MABS server in the Azure cloud to perform the same function as if on-premises.

Using a backup server, DPM, or MABS, will remove the need to have the MARS agent installed on each VM, as the MARS agent will only run on the backup server. However, each VM will be running the DPM or MABS protection agent. The backup server adds some features, more flexibility scheduling backups, app-aware backups, and provides a central place to manage multiple server backups that can be grouped into protection groups. This is a great option if there are multi-server applications (Web, Middleware, Backend DB). Grouping the servers

related to a single application into a protection group will allow you to back them up as one unit making restoration to a point in time much easier.

Backup retention and frequency are other important items to consider when developing a backup strategy. Note the following points:

Azure Backup Retention:

- Offers daily, weekly, monthly, and yearly retention, with up to 9999 recovery points.

- Backup frequency can impact the maximum retention period. A daily retention period will exhaust the 9999 max recovery points before meeting the monthly or yearly retention period.

Azure Backup Frequency Limitations:

- When backing up directly to the Azure Services Recovery Vault using the MARS agent, you can take three backups per day.

- Enabling backup on an Azure VM will allow only a weekly or daily backup. The backup servers, DPM, or MABS can be backed up two times per day.

- When backing up to disk (The DPM or MABS server), you can get up to every 15 minutes for SQL server and up to every hour for other workloads.

- Recovery points that are on the local disk of the backup server are limited to 64 for File Servers and 448 for Application Servers.

Security

Security of the data being backed up is another important consideration. Data should be encrypted while at rest, where the backup data is stored, and in transit, while the data is moving from one location to another. Azure backup can encrypt data both at rest and while in transit. While at rest, machines can use a customer-specified passphrase to encrypt data. Azure VMs use Storage Service

Encryption (SSE) on data stored in the vault. The backup process for Azure VMs automatically encrypts/decrypts data upon storage and retrieval, respectively.

You can back up Azure VMs that use Azure Disk Encryption (ADE); both BitLocker Encryption (BEK) and Key Encryption Key (KEK) are supported. Dm-crypt is also supported for Linux VMs.

When data is in transit, on-premises data is encrypted using AES256 and sent to the Azure vault over HTTPS. Data that originates in Azure and is moving to Azure Recovery Services vault is done over HTTPS and never leaves the Azure backbone network. When performing a restore within Azure, iSCSI secures the data, and secure tunneling protects the iSCSI channel.

The topic of security must include a discussion on ransomware.

Ransomware is malware that blocks access to data from the user or organization. It encrypts data and payment is demanded for the decryption key.

If an organization is affected by a ransomware attack, it is of utmost importance that the backed-up data is protected. Air Gapping the data on an isolated network segment as we discussed earlier helps protect the data. However, there is more that should be done. Making backed-up data immutable will prevent the backed-up data from being infected with malware that may have not been detected and backed up to the isolated network. At the time of writing Azure backup does not incorporate Airgap or Immutability in its solution.

By default, backup data traveling from on-premises to the Azure Recovery Services Vault will go over the internet using the HTTPS protocol for security. Alternately, a private endpoint can be configured on the vault to enable connectivity to a private Virtual Network (VNET). This will enable traffic to traverse an Express Route connection from an on-

premises datacenter to the Azure cloud. Express Route is a dedicated connection between a on premise datacenter and the Azure Cloud using an approved carrier.

Alternative Azure Backup Solutions

There are some key considerations when choosing a backup solution other than Azure, including:

Single User Interface (UI)

Our whole discussion on backup has been focused on Azure data protection options. And if we were looking at things myopically, that would be fine. Fact is, Azure is not the only cloud game in town. More and more companies utilize heterogenous hybrid environments and are leveraging multiple clouds, (AWS, Google, Azure, and others) to run their services. Also consider the fact that VMs and data are not the only resources that need to be backed up. With cloud adoption more vendors and organizations are re-factoring their applications to run on Platform as a Service (PaaS) solutions offered by the cloud vendors. There are also Software as a Service (SaaS) solutions such as Microsoft 365 that need to be considered.

The idea would be to simplify the backup and restoration of all your resources no matter which cloud they reside in or what type of resource they are. Using a cloud vendor specific solution would require you deploy and manage separate processes for each cloud. This is not a realistic approach when dealing with enterprise-level companies.

The best strategy is to have a solution that can be managed from one toolset with one management console regardless of where the data is located. A solution that can back up on-premises physical servers including SQL Server, Oracle, and SAP as well as VMware, Hyper-V, and other sources like Kubernates and NoSql, across different clouds such as AWS, GCP, and Azure. It should include PaaS and SaaS services as well. The ability to manage this through a single UI is a major consideration when seeking third-party assistance with data protection.

Hybrid Cloud Flexibility

Cloud adoption will continue to grow, but hybrid is the new normal most organizations have settled on. What this means is the need for flexibility and choice beyond a backup solution that is purpose

built for just one cloud or workload. First consider the need for a solution that supports a wide range of new cloud and legacy data sources as described above. Also, the need for flexibility on where to deploy your backups – in the cloud to protect cloud data sources and on-premises to protect the legacy sources – keep data close by for rapid recovery and reduce data transfer and data egress costs.

A choice of consumption models. Self-managed software and infrastructure for on-premises or if needed manage data in your own cloud accounts – or have it managed for you in a Backup as a Service (BaaS) model. But don't let flexibility and choice fool you into adding complexity and data silos with different tools and consoles to get the job done. Instead look for a solution that gives you hybrid cloud flexibility and choice that's built on the same platform and all managed through the same UI – no matter where you deploy and what consumption model you choose.

Leverage Backup Data (Dev/Test Use Cases)

We mentioned at the outset of our discussion, that backup is just a copy of data sitting at another

location just in case it is needed. Data just sitting and waiting, and waiting, and waiting. All this waiting for a disaster or corruption is costing money and depending on the amount of data it can be a significant cost. The ability to leverage that data and use it, while at the same time keeping it safe for a restore when needed, would really help make the cost more palatable. And finding a toolset that can do it would be ideal. This may even reduce costs in other areas. For example, what if you could use the backed-up data for testing purposes, bringing up a backed up VM to test an application upgrade or Operating System patch? This would negate the need for duplicate test servers. Or leverage the backed-up data to perform analytics that can drive Machine Learning and Artificial Intelligence.

Having solutions that can take on-premises servers and duplicate them in a cloud for dev/test purposes and easy tear-down is key for those who do a lot of dev/test work.

Compression and Deduplication

This is a no-brainer. The more you can deduplicate and compress data before you store it, the less cost you will have in storing your data. Azure backup agent MARS, DPM server, and MABS server will compress data before sending it to the Azure Recovery Services Vault. Azure does not deduplicate data.

A recent study by Cohesity, Inc found that 63% of organizations had between 4 and 15 copies of the same data.

Agent Management

Installing an agent on every protected machine adds to overall complexity and operational cost. These agents will need to be deployed, monitored, upgraded, etc. A backup solution that does not require an agent installed on every machine would simplify the solution. There are some vendors that use an appliance that can monitor for changed blocks on VMs (or, in the case of Azure, use disk

snapshots) and replicate these changes without the need for an agent on each VM.

Cybersecurity

There is a very high risk of an organization's data becoming compromised by a ransomware attack. When deciding on your backup strategy this must be considered. Implementation of software to prevent and detect any such attack would be the first line of defense. However, if your data does become compromised, the need for recovery of clean data, air-gapped and immutable, and the speed at which you can recover becomes the priority. The backup solution should allow you to recover from a point in time, should have the capability to ensure that the backed-up data has not been compromised, and should have the ability to restore the data in a short period of time.

The Big Takeaways

Azure Backup can be used to successfully back up both on-premises machines and Azure resources. The options available in Azure include Virtual Machines, Azure File Shares, SQL Server in Azure VM, and SAP HANA in Azure VM. On-premises options include Files and Folders, Hyper-V VMs, VMWare VMs, SQL Server, SharePoint, Exchange, System State, Bare Metal Recovery, and Azure Stack resources. When used with ASR, an acceptable BCDR strategy can be mapped out.

Azure Backup will compress, encrypt, and make redundant copies of backup data automatically. Azure will not deduplicate data. When using Azure Backup, there are some limits to what you can achieve depending on what type of server you are backing up. Using the MARS agent or DPM/MABS servers can add complexity to your solution. Azure backup will only work with the Azure Cloud. Many companies are adopting hybrid environments that leverage multiple cloud vendors, like AWS, Google, and others. Choosing a solution that can, not only, back up resources in multiple clouds, but can allow for management from a single UI will be important

for complete backup lifecycle success. Azure Backup does not allow access to backup data, except for restore. There are vendors that will allow you to make use of backup data resulting in better cost efficiencies. The fewer number of agents running on each server for a backup solution, the easier to manage, monitor, and upgrade the solution. Azure Backup on its own will not enable you to recover VMware servers in the Azure Cloud; you would need to add Azure Site Recovery ASR for this capability.

There are vendors that add additional security to backed up data. Along with encryption at rest and in transit. The added feature of making the backed-up data immutable further secures the data from becoming corrupt or accidentally deleted. Any new backup of the changed data is written to a separate disk, not even administrators have the ability to update/remove the data. Azure does not Airgap your backed up data, this is an important feature to consider when choosing a backup strategy. A solution that makes it easy to segment your backed-up data in and isolated network is imperative to a fully secure solution.

Backing up of data does not have to be so boring after all. It can be an exciting challenge to get the most efficient and reliable solution. There are many vendors offering cool and innovative technology in this space. Make sure you investigate well and come up with the best solution for the organization.

COHESITY

Azure Data is YOUR Responsibility

Simplify backups, instantly recover, and protect against Ransomware - with Cohesity.

https://www.cohesity.com/solutions/cloud/azure/

Whether you're operating in a single or multi-cloud environment, data protection is essential. Azure does include built-in tools, but they are limited. In this book we'll explore what these tools can and cannot do, and why you might seek out a third-party backup solution.

About Wayne Dipchan

Wayne Dipchan, a 20-year technical author, is a senior technology infrastructure architect for one of the largest health systems in the northeast US.