✓ 100 XP

Introduction

2 minutes

Classifying your data and identifying your data protection needs helps you select the right cloud solution for your organization. Data classification enables organizations to find storage optimizations that might not be possible when all data is assigned the same value. Classifying (or categorizing) stored data by sensitivity and business impact helps organizations determine the risks associated with the data. After your data has been classified, organizations can manage their data in ways that reflect their internal value instead of treating all data the same way.

Data classification can yield benefits such as compliance efficiencies, improved ways to manage the organization's resources, and facilitation of migration to the cloud. Some data protection solutions such as encryption, rights management, and data loss prevention have moved to the cloud and can help mitigate cloud risks. However, organizations must be sure to address data classification rules for data retention when moving to the cloud.

Learning objectives

In this module, you will:

- Learn how to classify your data
- Configure your data retention requirements
- Explore data ownership and sovereignty

✓ 100 XP

Classify your data at rest, in process, and in transit

8 minutes

Digital data always exists in one of three states: at rest, in process, and in transit.

All three states require unique technical solutions for data classification, but the applied principles of data classification should be the same for each. Data that is classified as **confidential** needs to stay confidential in each state.

Data can also be either **structured** or **unstructured**. Typical classification processes for structured data found in databases and spreadsheets are less complex and time-consuming to manage than those for unstructured data such as documents, source code, and email. Generally, organizations will have more unstructured data than structured data.

Regardless of the type of data, organizations need to manage data sensitivity. When properly implemented, data classification helps ensure that sensitive or confidential data assets are managed with greater oversight than data assets that are considered public distribution.

Protect data at rest

Data encryption at rest is a mandatory step toward data privacy, compliance, and data sovereignty.

Best practice	Solution
Apply disk encryption to help safeguard your data.	Use Microsoft Azure Disk Encryption, which enables IT administrators to encrypt both Windows infrastructure as a service (laaS) and Linux laaS virtual machine (VM) disks. Disk encryption combines the industry-standard BitLocker feature and the Linux DM-Crypt feature to provide volume encryption for the operating system (OS) and the data disks. Azure Storage and Azure SQL Database encrypt data at rest by default,

Best practice	Solution
	and many services offer encryption as an option. You can use Azure Key Vault to maintain control of keys that access and encrypt your data. See Azure resource providers encryption model support to learn more.
Use encryption to help mitigate risks related to unauthorized data	Encrypt your drives before you write sensitive data to them.
access.	

Organizations that don't enforce data encryption risk higher exposure to data-integrity issues. For example, unauthorized users or malicious hackers might steal data in compromised accounts or gain unauthorized access to data coded in Clear Format. To comply with industry regulations, companies also must prove that they are diligent and using correct security controls to enhance their data security.

Protect data in transit

Protecting data in transit should be an essential part of your data protection strategy. Because data is moving back and forth from many locations, we generally recommend that you always use SSL/TLS protocols to exchange data across different locations. In some circumstances, you might want to isolate the entire communication channel between your on-premises and cloud infrastructures by using a VPN.

For data moving between your on-premises infrastructure and Azure, consider appropriate safeguards such as HTTPS or VPN. When sending encrypted traffic between an Azure virtual network and an on-premises location over the public internet, use Azure VPN Gateway.

The following table lists best practices specific to using Azure VPN Gateway, SSL/TLS, and HTTPS.

Best practice	Solution
Secure access from multiple workstations located on-premises to an Azure virtual network	Use site-to-site VPN.

Best practice	Solution
Secure access from an individual workstation located on-premises to an Azure virtual network	Use point-to-site VPN.
Move large data sets over a dedicated high-speed wide-area network (WAN) link	Use Azure ExpressRoute. If you choose to use ExpressRoute, you can also encrypt the data at the application level by using SSL/TLS or other protocols for added protection.
Interact with Azure Storage through the Azure portal	All transactions occur via HTTPS. You can also use Storage REST API over HTTPS to interact with Azure Storage and Azure SQL Database.

Organizations that fail to protect data in transit are more susceptible to man-in-the-middle attacks, eavesdropping, and session hijacking. These attacks can be the first step in gaining access to confidential data.

Data discovery

Data discovery and classification (currently in preview) provides advanced capabilities built into Azure SQL Database for discovering, classifying, labeling and protecting sensitive data (such as business, personal data (PII), and financial information) in your databases. Finding and classifying this data can play a pivotal role in your organizational information protection stature. It can serve as infrastructure for:

- Helping meet data privacy standards and regulatory compliance requirements.
- Addressing various security scenarios such as monitoring, auditing, and alerting on anomalous access to sensitive data.
- Controlling access to and hardening the security of databases containing highly sensitive data.

Data discovery and classification is part of the <u>Advanced Data Security</u> offering, which is a unified package for advanced Microsoft SQL Server security capabilities. You access and manage data discovery and classification via the central SQL Advanced Data Security portal.

Data discovery and classification introduces a set of advanced services and SQL capabilities, forming a SQL Information Protection paradigm aimed at protecting the data, not just the database:

- **Discovery and recommendations** The classification engine scans your database and identifies columns containing potentially sensitive data. It then provides you with a more natural way to review and apply the appropriate classification recommendations via the Azure portal.
- Labeling Sensitivity classification labels can be persistently tagged on columns
 using new classification metadata attributes introduced into the SQL Server
 Engine. This metadata can then be utilized for advanced sensitivity-based auditing
 and protection scenarios.
- **Query result set sensitivity** The sensitivity of the query result set is calculated in real-time for auditing purposes.
- Visibility You can view the database classification state in a detailed dashboard in the Azure portal. Additionally, you can download a report (in Microsoft Excel format) that you can use for compliance and auditing purposes, in addition to other needs.

Steps for discovery, classification, and labeling

Classifications have two metadata attributes:

- **Labels** These are the main classification attributes used to define the sensitivity level of the data stored in the column.
- Information Types These provide additional granularity into the type of data stored in the column.

SQL data discovery and classification comes with a built-in set of sensitivity labels and information types and discovery logic. You can now customize this taxonomy and define a set and ranking of classification constructs specifically for your environment.

Customization of your classification taxonomy takes place in one central location for your entire Azure tenant: **Azure Security Center**. Only a user with administrative rights on the Azure tenant root management group can perform this task.

As part of Azure Information Protection policy management, you can define custom labels, rank them, and associate them with a selected set of information types. You can also add your own custom information types and configure them with string patterns, which are added to the discovery logic for identifying this type of data in your databases. Learn more about customizing and managing your policy with the links in the Summary of this module.

After you've defined the tenant-wide policy, you can continue with classifying individual databases using your customized policy. Let's examine this in more detail with Azure SQL DB.

Next unit: Exercise - Classify an Azure SQL Database

Continue >

✓ 100 XP

Exercise - Classify an Azure SQL Database

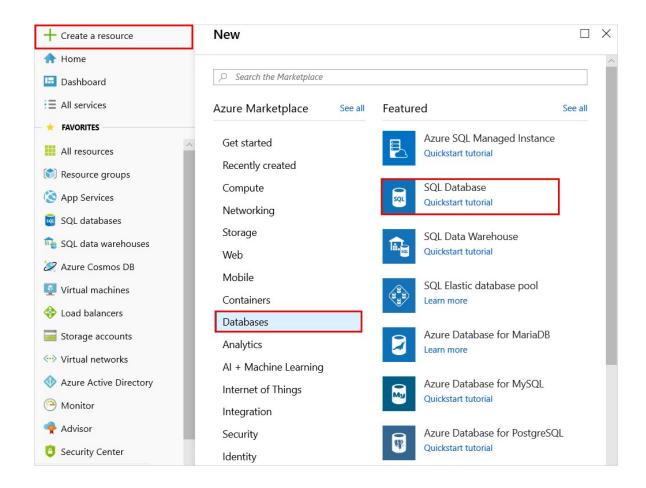
10 minutes

This module requires a sandbox to complete. A <u>sandbox</u> gives you access to Azure resources. Your Azure subscription will not be charged. The sandbox may only be used to complete training on Microsoft Learn. Use for any other reason is prohibited, and may result in permanent loss of access to the sandbox.

Activate sandbox

In this step, you will create your resource group and an Azure SQL Database single database containing the AdventureWorksLT sample data

- 1. Sign-in to the <u>Azure portal</u> □ using the same account you used to activate the Azure Sandbox. Make sure you are in the Microsoft Learn Sandbox directory.
- 2. Select + Create a resource in the left sidebar of the Azure portal.
- Select Databases and then select SQL Database to open the Create SQL Database page.



4. Use these values to fill out the form.

Setting	Value
Database name	LearnDataPolicy
Subscription	Concierge Subscription
Resource group	Use the existing group [sandbox resource group name]
Want to use SQL elastic pool?	No

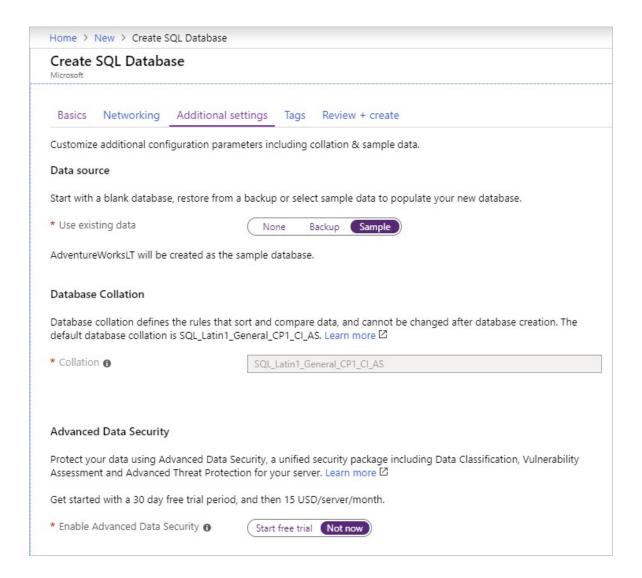
5. Under **Server**, click **Create new**, fill out the form, then click **OK**. Here's more information on how to fill out the form:

Setting	Value
---------	-------

Setting	Value
Server	
name	
Server	A database identifier that serves as your primary administrator login
admin login	name.
Password	Any valid password that has at least eight characters and contains
	characters from three of these categories: uppercase characters,
	lowercase characters, numbers, and non-alphanumeric characters.
Location	Any valid location from the available list below.

The free sandbox allows you to create resources in a subset of the Azure global regions. Select a region from the following list when you create resources:

- West US 2
- South Central US
- Central US
- East US
- West Europe
- Southeast Asia
- Japan East
- Brazil South
- Australia Southeast
- Central India
- 6. Select the **Additional settings** tab.
- 7. In the **Data source** section, under **Use existing data**, select **Sample**.



- 8. Under Enable Advanced Data Security, verify Start free trial is selected.
- 9. Leave the rest of the values as default and select **Review + Create** at the bottom of the form.
- 10. Review the final settings and select **Create**.

It will take a few minutes to deploy the server with sample data. Once it's complete, select **Go to resource** to navigate to the Overview view of your new SQL database.

SQL Information Protection (SQL IP)

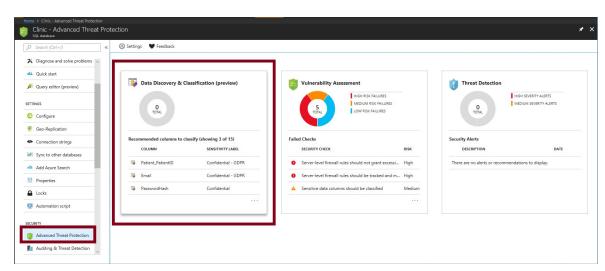
SQL IP brings a set of advanced services and SQL capabilities, forming a new information protection paradigm in SQL aimed at protecting the data, not just the database:

- **Discovery & recommendations** The classification engine scans your database and identifies columns containing potentially sensitive data. It then provides you an easy way to review and apply the appropriate classification recommendations via the Azure portal.
- Labeling Sensitivity classification labels can be persistently tagged on columns
 using new classification metadata attributes introduced into the SQL Engine. This
 metadata can then be utilized for advanced sensitivity-based auditing and
 protection scenarios.
- **Monitoring/Auditing** Sensitivity of the query result set is calculated in real time and used for auditing access to sensitive data (currently in Azure SQL DB only).
- Visibility The database classification state can be viewed in a detailed dashboard in the portal. Additionally, you can download a report (in Excel format) to be used for compliance & auditing purposes, as well as other needs.

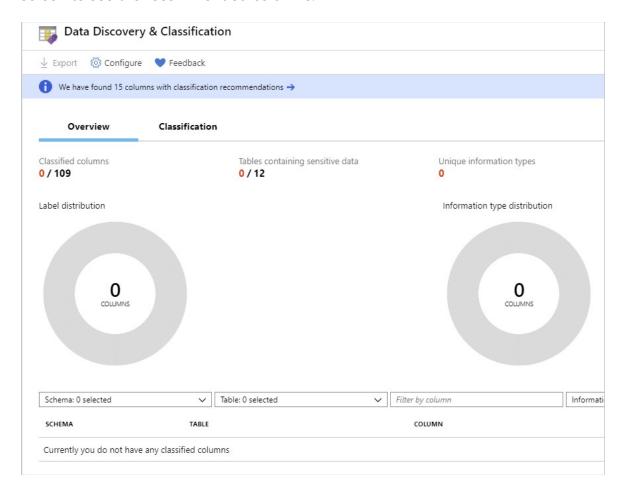
Classify your SQL DB

Let's classify the data in this sample Azure SQL database.

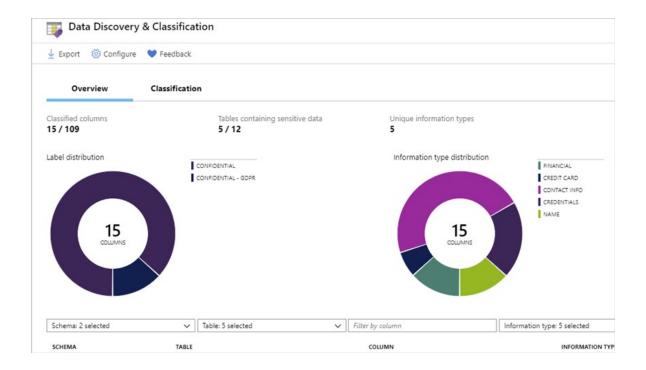
- 1. Under the **Security** heading in the Azure SQL Database pane, navigate to **Advanced Data Security**.
- 2. If Advanced Data Security isn't enabled, select the **Enable** button to enable it. As noted in the instructions above, you can turn this on as part of the DB creation. This will take a minute to activate.
- 3. Select the **Data Discovery and Classification** card.



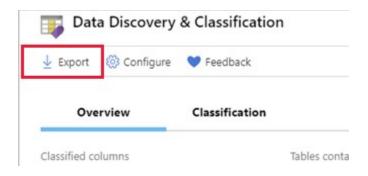
4. The initial state will have recommended columns - 15 in this case, but none of them will be classified yet. Select the recommendations info tip at the top of the screen to see the recommended columns.



- 5. Select all the columns and then **Accept selected recommendations**.
- 6. Select **Save** to save the recommendations, and then switch back to the **Overview** tab.
- 7. Review the **Overview** tab. Notice that it includes a summary of the current classification state of the database, including a detailed list of all classified columns. You can also filter this view to only see specific schema parts, information types, and labels.



8. To download a report in Excel format, in the top menu of the window select Export.



Customizing the classification

The **Classification** tab lists the columns and how the data is classified.



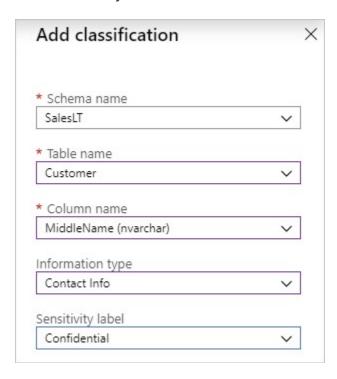
As you saw earlier, the classification engine scans your database for columns containing potentially sensitive data and provides a list of recommended column classifications.

You can either take the suggested classifications as we did earlier, or manually classify columns as an alternative to or in addition to the recommendation-based classification.

1. In the top menu of the window, select **Add classification**.



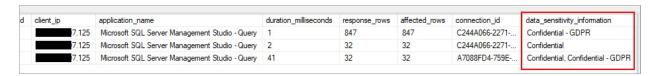
- 2. In the Add classification pane, configure the five fields that display, and then select Add classification:
 - Schema name
 - Table name
 - Column name
 - Information type
 - Sensitivity label.



- 3. To complete your classification and persistently label (tag) the database columns with the new classification metadata, in the top menu of the window, select **Save**.
- 4. Switch back to the **Overview** tab to now see **16** columns identified.
- 5. Try changing some of the other classifications for identified columns for example, setting the password data to **Highly Confidential**.
- 6. You can also filter the data being viewed through the filter boxes right below the graphs on the **Overview** tab.

Monitor access to sensitive data

An important aspect of the IP paradigm is the ability to monitor access to sensitive data. <u>Azure SQL Database Auditing</u> has been enhanced to include a new field in the audit log. The data_sensitivity_information field logs the sensitivity classifications (labels) of the actual data that was returned by the query.



Consider configuring Azure SQL Database Auditing for monitoring and auditing access to your classified sensitive data.

Next unit: Explore data recovery, retention, and disposal

Continue >

✓ 100 XP

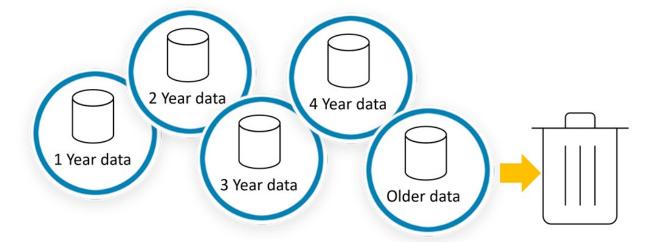
Explore data recovery, retention, and disposal

9 minutes

Once an organization's data has been examined and classified, the next decision to make is how long to keep the data around. Data recovery and disposal is an essential aspect of managing data assets. A *data retention policy* defines the principles for data recovery and disposal and enforced in the same manner as data reclassification. These tasks are typically performed by the custodian and administrator roles as a collaborative task.

Failure to maintain a data retention policy could mean data loss or failure to comply with regulatory and legal discovery requirements. Most organizations that do not have a clearly defined data retention policy tend to use a default, Keep everything retention policy. However, this poses additional risks in cloud services scenarios. For example, a data retention policy for cloud service providers can be considered as "for the duration of the subscription," meaning as long as the service is paid for, the data is retained. Such a pay-for-retention agreement might not address corporate or regulatory retention policies.

Defining a policy for confidential data can ensure that data is stored and removed based on best practices. Also, you can create an archival policy to formalize an understanding of what data should be disposed of and when.



A data retention policy should address the required regulatory and compliance requirements and corporate legal retention requirements. Properly classified data should influence decisions made about retention duration. Data classification rules that pertain to data retention must be addressed when moving to the cloud. Data protection technologies such as encryption, rights management, and data loss prevention solutions are available in the cloud and can help mitigate disclosure risks.

Immutable storage and data retention

Immutable storage for Azure Blob Storage enables users to store business-critical data in a write once, read many (WORM) state. This state makes the data unerasable and unmodifiable for a user-specified interval. Blobs can be created and read, but not modified or deleted, for the duration of the retention interval.

Immutable storage enables:

- Time-based retention policy support Users set policies to store data for a specified interval.
- **Legal hold policy support** When the retention interval is not known, users can set legal holds to store data immutably until the legal hold is cleared. When a legal hold is set, blobs can be created and read, but not modified or deleted. Each legal hold is associated with a user-defined alphanumeric tag that is used as an identifier string (such as a case ID).
- **Support for all blob tiers** WORM policies are independent of the Azure Blob Storage tier and apply to all tiers: hot, cool, and archive. Users can transition data

to the most cost-optimized tier for their workloads while maintaining data immutability.

- **Container-level configuration** Users can configure time-based retention policies and legal hold tags at the container level. By using simple container-level settings, users can create and lock time-based retention policies, extend retention intervals, set and clear legal holds, and more. These policies apply to all the blobs in the container, both existing and new.
- Audit logging support Each container includes an audit log, which displays up
 to five time-based retention commands for locked time-based retention policies.
 However, the log has a maximum of three logs for retention interval extensions or
 time-based retention. The log contains the user ID, command type, time stamps,
 and retention interval. For legal holds, the log contains the user ID, command type,
 time stamps, and legal hold tags.

The audit log is kept for the lifetime of the container, in accordance with the SEC 17a-4(f) regulatory guidelines. The Azure Activity Log shows a more comprehensive log of all the control plane activities. It is the user's responsibility to store those logs persistently, as might be required for regulatory or other purposes.

Immutable storage for Azure Blob storage supports two types of WORM or immutable policies: time-based retention and legal holds.

When a time-based retention policy or a legal hold is applied on a container, all existing blobs move to the immutable (write-protected and delete-protected) state. All new blobs that are uploaded to the container will also move to the immutable state.

When a time-based retention policy is applied on a container, all blobs in the container will stay in the immutable state for the duration of the effective retention period. The effective retention period for existing blobs is equal to the difference between the blob creation time and the user-specified retention interval.

For new blobs, the effective retention period is equal to the user-specified retention interval. Because users can extend the retention interval, immutable storage uses the most recent value of the user-specified retention interval to calculate the effective retention period.

For example, a user creates a time-based retention policy with a retention interval of five years. The existing blob in that container, testblob1, was created one year ago. The

effective retention period for testblob1 is four years. A new blob, testblob2, is now uploaded to the container. The retention period for this new blob is five years.

Legal holds

When a reasonable expectation of litigation exists, organizations are required to preserve electronically stored information (ESI). This expectation often exists before the specifics of the case are known, and preservation is often broad. When you set a legal hold, all new and existing blobs stay in the immutable state until the legal hold is cleared.

A container can have both a legal hold and a time-based retention policy simultaneously. All blobs in that container stay in the immutable state until all legal holds are cleared, even if their effective retention period has expired. Conversely, a blob stays in an immutable state until the effective retention period expires, even though all legal holds have been cleared.

Next unit: Understand data sovereignty

Continue >

✓ 100 XP

Understand data sovereignty

8 minutes

Digital information is always subject to the laws of the country or region where it's stored. This concept is known as *data sovereignty*. Many of the concerns that surround data sovereignty relate to enforcing privacy regulations and preventing data that are stored in a foreign country from being subpoenaed by the host country or region's government.

In Azure, customer data can be replicated within a geographic area for enhanced data durability if there's a significant data center disaster.

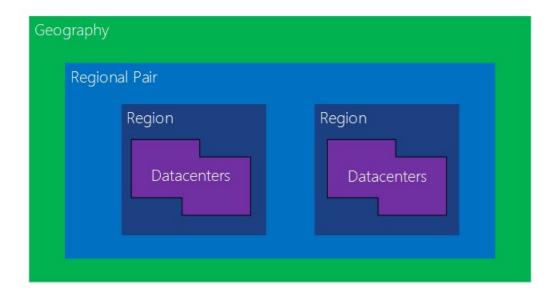
(i) Important

No matter where customer data is stored, Microsoft does not control or limit the locations from which customers or their end users might access their data.

Paired regions

Azure operates in multiple geographies around the world. Azure geography is a defined area of the world that contains at least one **Azure Region**. An Azure region is an area containing one or more data centers.

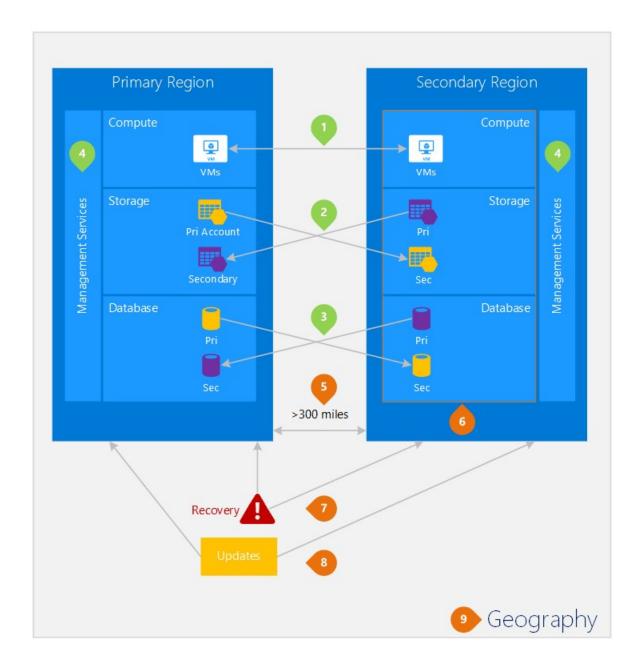
Each Azure region is paired with another region within the same geography, forming a *region pair*. The exception is Brazil South, which is paired with a region outside its geography. Across the region pairs Azure serializes platform updates (or planned maintenance) so that only one region is updated at a time. If an outage affecting multiple regions occurs, one region in each pair will be prioritized for recovery.



It's recommended that you configure business continuity and disaster recovery (BCDR) across regional pairs to benefit from Azure's isolation and VM policies. For applications that support multiple active regions, we recommend using both regions in a region pair where possible. This approach will ensure optimal availability for applications and minimized recovery time in the event of a disaster.

An example of paired regions

The following illustration is of a hypothetical application that uses a regional pair for disaster recovery. The green numbers highlight the cross-region activities of three Azure services (Azure Compute, Azure Storage, and Azure Database) and how they are configured to replicate across regions. The orange numbers highlight the unique benefits of deploying across paired regions.



Cross-region activities number key

- **Azure Compute (laaS)** You must provision additional compute resources in advance to ensure resources are available in another region during a disaster. For more information, see Designing resilient applications for Azure.
- Azure Storage- Geo-redundant storage (GRS) is configured by default when an
 Azure Storage account is created. With GRS, data is automatically replicated three
 times within the primary region, and three times in a paired region. For more
 information, see Azure Storage redundancy.

- Azure SQL Database With Azure SQL Database geo-replication, you can
 configure asynchronous replication of transactions to any region in the world;
 however, we recommend you deploy these resources in a paired region for most
 disaster recovery scenarios. For more information, see Configure active georeplication for Azure SQL Database in the Azure portal.
- **Azure Resource Manager** Resource Manager inherently provides logical isolation of components across regions. This means that logical failures in one region are less likely to impact other regions.

Benefits of Azure paired regions number key:

- Physical isolation When possible, Azure services prefer at least 300 miles of separation between datacenters in a regional pair (although this isn't practical or possible in all geographies). Physical datacenter separation reduces the likelihood of both regions being affected simultaneously as a result of natural disasters, civil unrest, power outages, or physical network outages. Isolation is subject to the constraints within the geography, such as geography size, power, and network infrastructure availability, and regulations.
- **Platform-provided replication** Some services such as geo-redundant storage provide automatic replication to the paired region.
- Region recovery order In the event of a widespread outage, recovery of one
 region is prioritized out of every pair. Applications that are deployed across paired
 regions are guaranteed to have one of the regions recovered with priority. If an
 application is deployed across regions that are not paired, recovery might be
 delayed. In the worst case, the chosen regions might be the last two to be
 recovered.
- **Sequential updates** Planned Azure system updates are rolled out to paired regions sequentially, not at the same time. This helps minimize downtime, the effect of bugs, and logical failures in the rare event of a bad update.
- **Data residency** To meet data residency requirements for tax and law enforcement jurisdiction purposes, a region resides within the same geography as its pair (with the exception of Brazil South).

Microsoft also complies with international data protection laws regarding transfers of customer data across borders. For example, to accommodate the continuous flow of information required by the international business (including the cross-border transfer

of personal data), many Microsoft business cloud services offer customers <u>European</u> <u>Union Model Clauses</u> that provide additional contractual guarantees around transfers of personal data for in-scope cloud services. European Union data protection authorities have validated the Microsoft implementation of the EU Model Clauses as being in line with the rigorous privacy standards that regulate international data transfers by companies operating in its member states.

In addition to our commitments under the Standard Contractual Clauses and other model contracts, Microsoft is certified to the EU-U.S. Privacy Shield Framework as set forth by the U.S. Department of Commerce regarding the collection, use, and retention of personal information transferred from the European Union to the United States. Microsoft participation in the EU-U.S. Privacy Shield applies to all personal data that is subject to the Microsoft Privacy Statement and is received from the EU, European Economic Area, and Switzerland. Microsoft also abides by Swiss data protection law regarding the processing of personal data from the European Economic Area and Switzerland.

(i) Important

Microsoft will not transfer to any third party (not even for storage purposes) data that you provide to Microsoft through the use of our business cloud services, and that are covered under the <u>Microsoft Online Services Terms</u>.

Next unit: Summary

Continue >

2 minutes

Summary

Data classification and analysis helps you know what data you need to protect and assess the risks involved in using, storing, and transmitting that data. Every organization has unique data that has specific security requirements. Unless specific work is done to identify the risks involved in a security breach, the costs and tools needed to protect that data can't be fully understood.

Clean up

The sandbox automatically cleans up your resources when you're finished with this module.

When you're working in your own subscription, it's a good idea at the end of a project to identify whether you still need the resources you created. Resources left running can cost you money. You can delete resources individually or delete the resource group to delete the entire set of resources.

In this module we turned on **Advanced Data Security** on an Azure SQL database. This has a monthly cost above and beyond the DB storage itself. If you did this work in your own subscription, make sure to turn this feature off to avoid unexpected costs.

Further reading

To learn more about the topics covered in this module, read through the following resources.

- Azure Security Center
- Azure Information Protection
- Azure SQL Database and SQL Data Warehouse data discovery & classification
- Information protection in Microsoft 365

Module complete:

Unlock achievement