Microsoft

Seven Key Principles of Cloud Security and Privacy

Mitigate vulnerabilities and protect your cloud data warehouse

The essential principles of security and privacy for your cloud data warehouse

by Donald Farmer
Principal
TreeHive Strategy

Contents

Principles of security and privacy for the cloud data warehouse

Almost every day we read of data breaches, the egregious misuse of customer information or some new hacking scandal. In this treacherous environment, any company using data extensively (and don't we all?) must grapple with ever more regulation and acute consumer pressure. As a result, security and privacy of data are top of mind for both anxious executives and hard-pressed IT teams.

These demands prove particularly troubling for data warehouse architects and administrators. The data warehouse, by design, models and stores notably high-value data, integrated, cleansed and denormalised to enable efficient analysis. A valuable source of insight for business, it can also be a gold mine for intruders if not rigorously designed for security and privacy, and then sedulously protected.

However, despite the sobering public scrutiny of our failures and broad discussion of practices, IT architects and administrators often find the demands of compliance and trust tricky to align with specific features of their chosen software. They face this difficulty especially when services are implemented and configured, as they often are today, in a complex framework which embraces cloud platforms, hybrid data management platforms and client applications.

Indeed, the Azure SQL Data Warehouse platform offers by far the most comprehensive set of compliance and security capabilities of any cloud data warehouse provider.

Microsoft has deep involvement in all these areas and global teams who advise both internal product teams and customers or partners on compliance. Indeed, the **Azure SQL Data Warehouse** platform offers by far the most comprehensive set of compliance and security capabilities of any cloud data warehouse provider.

This paper addresses seven key principles and practices building on this hard-won experience. The principles are common to all cloud data warehousing scenarios. The practices described here are specific to the **Azure SQL Data Warehouse**; surprisingly many of them are poorly supported on other platforms. Understanding these fundamentals will help you to refine your own policies and processes. Consequently, without distracting from the complexity of some scenarios, you will be better prepared to implement and support the most demanding privacy and security initiatives today and in the future, as requirements evolve in the new business and regulatory landscape.

Introduction: The new business environment of security and privacy

As you follow the news, you can sense an ever-growing gap between the reality of consumers' demands and the reassuring myths of security that many software vendors tell.

Consumers are certainly becoming more aware, and more concerned, about the privacy and security of their data, especially with cloud services and storage. Software vendors and service providers, on the whole, work very hard to address these concerns. Their first motivation may well be the demands of new regulations and the possibility of stringent sanctions. However, along with the legal stick, we should consider the carrot of competitive advantage. A vendor who respects the privacy and consent of users can win new customers and retain an existing community with greater loyalty than before. Well-secured data also enables operational efficiencies within the enterprise, from on-boarding new users to provisioning data and even enabling a more data-driven partner ecosystem. In short, privacy and security are not a burden for you to carry but a commitment that benefits you and your customers.

Regulation

There have been numerous attempts to legislate better protections for consumer data. Some of these efforts, such as the British Standard 7799 and the US Health Insurance Portability and Accountability Act (HIPPA) have been around since the 1990s and are well established.

Today, many discussions of security and privacy legislation start with the European Union's General Data Protection Regulation (GDPR) which came into effect in May 2018. The GDPR aims to systematise data privacy laws across Europe, to protect the personal data of EU citizens, giving them control over the processing and use of that data and to reform the approach to data privacy of organisations across the region. With 99 articles and 171 explanatory recitals, the GDPR may be unusually comprehensive, but its intentions are not extraordinary. The regulation has served as a model for other efforts worldwide, for example:

- Notifiable Data Breaches scheme, Australia
- Personal Information Protection Act, Japan
- New York State Department of Financial Services cybersecurity regulation 23 NYCRR 500
- California Consumer Privacy Act of 2018
- Colorado's Consumer Data Protection Laws

In particular, four aspects of the GDPR's approach have been influential.

- Enterprises must receive explicit consent from consumers to use their data in specific ways.
- Consumers have the right to access their data including the right to demand prompt corrections or deletion of data.
- Controllers of data must put into practice measures which meet the principles of data protection by design and by default.
- The regulations are risk-based; that is to say, rather than following a strict set of procedures defined by the regulators, data controllers must instead evaluate the likelihood and severity of incidents, then take appropriate steps to mitigate the danger and minimise impact.

It can be difficult to keep up with the changing landscape of regulation. The Microsoft Service Trust Portal provides a variety of very useful content, tools and other resources about Microsoft security, privacy and compliance practices.

Consent and the right to access have important implications for user agreements, terms of service and even the user interface design of sign-up screens and forms. This paper is most concerned with the measures that must be implemented by design and by default, and how those same measures can influence security work.

The need for principles

There are numerous regulations governing data privacy and security, applicable at regional, national and international levels. The **Microsoft Service Trust Portal** is a very useful resource for enterprises working to meet the requirements of these rules.

This paper will identify some basic principles underlying both legal and commercial pressures to help you define your own strategic approach, rather than detailing the specifics of each regulation.

Principle 1: Security and privacy are significantly different, but closely related

To be free from threat and unwanted attention or intrusion are basic human needs which, nevertheless, vary widely in their terms between different societies and social contexts.

Even when discussing data security and privacy in the context of regulations, there are important, but subtle, differences between different regimes. So, the specific definitions that apply in your circumstances will vary, but it is still useful to have some fundamental principles in mind to guide your thinking.

If we take privacy to mean that freedom from unwanted attention or intrusion mentioned earlier, and security as the need to be free from threat, then a simple analogy may make the important differences clear.

In your home, if your blinds are open and anyone can look in, you have little privacy, even though, with the doors and windows locked, you feel secure. However, if your blinds are closed but the door unlocked, despite your sense of privacy, which no passer-by can casually breach, you are not secure. And indeed, if someone enters through your unlocked door, your privacy as well as your security have been compromised.

In terms of data, think of the controls you have in place to ensure the security of your network, such as firewalls and logins. You may be secure from unauthorised access, but if you have not taken care to ensure that even authenticated users do not casually see what they are not meant to see, you are not protecting privacy.

On the other hand, perhaps you have put in place protections to ensure that only users with the right permissions see certain data. You have done your best to protect privacy in that context. Yet, if the network is vulnerable, unauthorised users may still get access – if they acquire the right privileges, both security and privacy are breached.

It is very important to understand, in any one scenario, what you are protecting: security or privacy? The answer is often both, but you will still find it best to consider them as separate scenarios.

Practices

With an understanding that security and privacy are separate but associated subject areas, you will find it practical to establish two separate but associated programmes. A smaller, virtual team can co-ordinate between the initiatives.

With an understanding that security and privacy are separate but associated subject areas, you will find it practical to establish two separate but associated programmes.

The focus of a privacy programme should be on the appropriate legislation that determines processing, protection and retention requirements. The programme must also take into account consumer expectations. Realistically, compliance with regulations will be top of the agenda but do remember that a privacy programme that meets or exceeds consumer expectations may be a competitive advantage.

Given the legal requirements, in addition to the technical members of a typical data governance team, you will also need a legal specialist on the privacy team. Given the commercial advantage of getting this right, you will likely need someone from marketing to be involved too.

The requirements identified by the privacy programme are passed to the security programme for implementation. However, the privacy team will not decide the specific processes or technology to be implemented, although they will have well-informed suggestions.

It is the security team that has the knowledge and capability to set up suitable protections and controls to meet the needs identified by the privacy team. An important advantage of keeping the two separated in this way is that the interpretation of regulations will not be unduly influenced by concerns about technical limitations. As a result, when the security team reviews the unbiased requirements from the privacy programme, it is possible they will identify technical demands which require new, or different, tools and platforms. This leads us to the second principle...

Principle 2: Security and privacy start with the platform

The days when data warehousing teams primarily governed a three-tier architecture on premises are long past. Today, users access analytic systems from browser-based tools, richly-featured desktop applications, mobile apps or, increasingly, through APIs. Data storage and the logical model may be in the cloud, on-premises or in a hybrid architecture which itself can vary from region to region depending on local requirements, regulations and available bandwidth. The components of an analytic system rarely come from a single vendor. There is no 'one throat to choke' anymore and trying to insist on a centrally-standardised architecture is largely futile: users adopt self-service for their favourite BI applications, data scientists increasingly focus on techniques from open source libraries and even IT employs a varied portfolio of tools.

The challenge here is not just to implement effective security and privacy processes. There is also the risk of deploying unnecessary, cumbersome or inconsistent controls if you apply separate security policies in each component of your infrastructure. The fundamentals of a privacy programme need to apply with an evenness of attention across all the components of your portfolio.

The more support the platform provides, the less you have to worry about and the more consistently you can apply your best practices.

Practices

In general, for any scenario, it is a sound practice to define and implement security and privacy controls at the lowest practical tier – as close to your data storage as possible. Benefits will bubble up from there.

On the cloud, however, there is another consideration. You should look to the underlying cloud platform itself to provide excellent security and privacy features. The more support the platform provides, the less you have to worry about and the more consistently you can apply your best practices.

Encryption, for example, is a critical feature for both security and privacy. Look for comprehensive encryption features in your storage platform. At the least, expect to protect your data with both encryption-at-rest and encryption-in-flight.

In the **Azure SQL Data Warehouse**, encryption-at-rest is enabled by **Transparent Data Encryption**: TDE. Once enabled by the administrator, at the server level, TDE performs real-time encryption and decryption of the database, associated backups and transaction log files at rest, without requiring changes to the logical model or client applications.

Transparent Data Encryption uses a symmetric key – that is to say, the same key is used for both encryption and decryption. This database encryption key is itself safeguarded by the TDE protector. The protector is either a built-in server certificate (service-managed transparent data encryption) or an asymmetric key stored in **Azure Key Vault** (Bring Your Own Key).

Some enterprises, by policy, wish to retain control and management of their encryption keys. For this scenario, the TDE protector is stored in a customer-owned and managed Azure Key Vault. The Key Vault administrator grants access to the logical SQL database server level and is inherited by all databases associated with that server. The Key Vault administrator may revoke access to the keys at any time.

However, there are advantages to using the service-managed certificate. If two databases are connected to the same server, they share the same built-in certificate. Microsoft automatically rotates these certificates in compliance with the internal security policy and the root key is protected by a Microsoft internal secret store. Microsoft also moves and manages the keys as needed for geo-replication and restores.

Encryption-in-flight is enabled by the industry-standard Transport Layer Security protocol: TLS. This protocol encrypts data in transit to and from the database, protecting from man-in-the-middle attacks. Your best practice will be to always use connections secured in this way and ensure that TLS 1.2 is supported as this is the recommended protocol to use for highly secure communication.

You will likely spend a good deal of time and effort defining and implementing your data privacy and security policies. So it's immensely frustrating when, despite your own best efforts, the platform itself lets you down, despite the gilt-edged claims of vendors that they put security first. ***Microsoft SQL Server – the core technology of Azure SQL Data Warehouse – has been the least vulnerable database over the last 8 years in the NIST vulnerabilities database***.[1]

Principle 3: We are all potential targets

Each year, **Verizon** publishes a **Data Breach Investigations Report** analysing thousands of real-world incidents. In 2018 they looked at over 53,000 incidents, including over 2000 confirmed data breaches. Their conclusion from studying this sample? *It will probably be you one day*.[2]

Perhaps even more concerning is that organisations are unable to identify suspicious database activities in good time. Nearly 70% of security events were reported only several months after they had occurred![2]

For anyone working to meet the requirements of new regulations, these are worrying statistics. Most regulations not only demand that you protect your systems with best practices, they also require timely reporting of data breaches.

For data warehouse architects, the situation is perhaps even more concerning. By their nature, analytic systems tend to include high-value data, which has often been cleansed, integrated and denormalised specifically to make analysis efficient and simple. But now any data breach may have access to information which is easier to exploit than more atomic operational data.

In 2018 they looked at over 53,000 incidents, including over 2000 confirmed data breaches. Their conclusion from studying this sample? It will probably be you one day.[2]

[1] https://nvd.nist.gov/vuln

[2] https://enterprise.verizon.com/resources/reports/DBIR_2018_Report_execsummary.pdf

Practices

In order to confront these threats with confidence, we need to put two practices in place. Firstly, regularly testing our systems for vulnerabilities and then continuously analysing our systems for anomalous behaviours that often indicate a threat as it occurs.

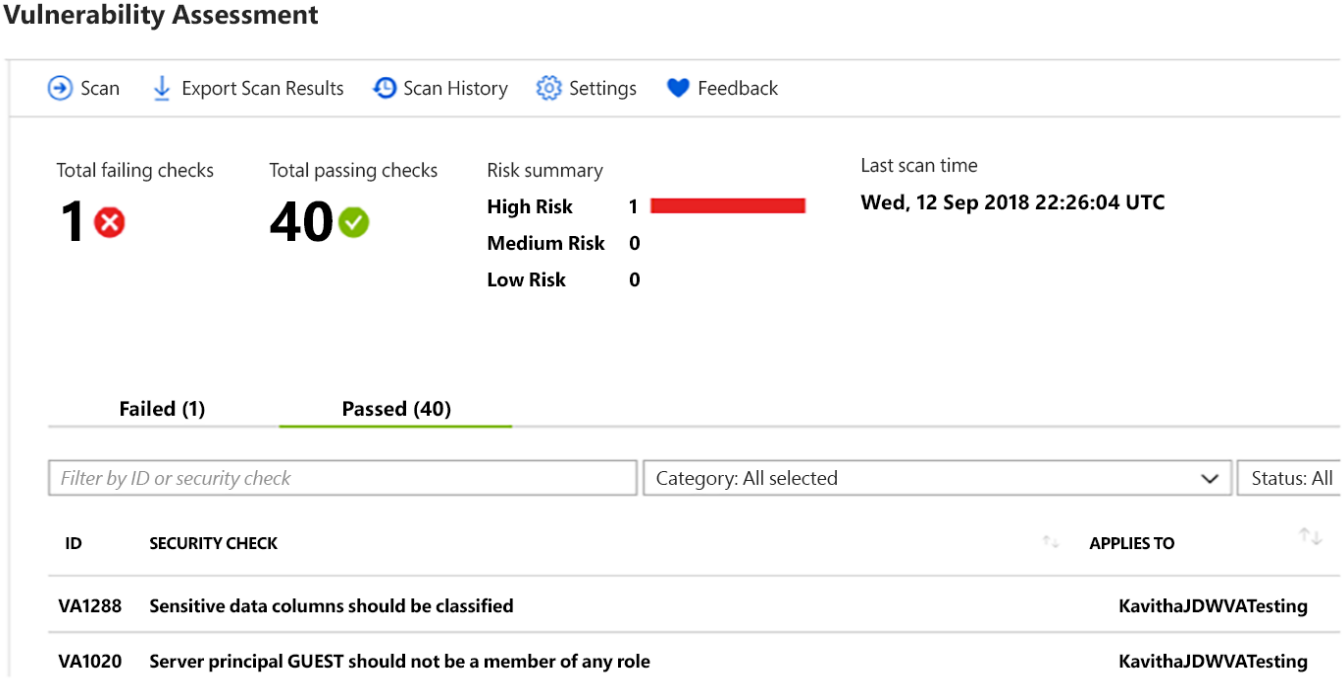
Microsoft Azure SQL Data Warehouse has features and services specifically designed for these needs.

SQL Vulnerability Assessment

The **SQL Vulnerability Assessment** is a scanning service, built into **Azure SQL Data Warehouse**, which evaluates and documents your system security and recommends steps to resolve security issues. The service also helps you to keep your system under systematic review, which is especially useful for a data warehouse environment where change is constant and difficult to keep tabs on.

To achieve this, the service uses a knowledge base of rules based on Microsoft's best practices that identify security vulnerabilities and lapses. For example, the service may find database-level issues such as excessive permissions or unprotected sensitive data. But the rules also identify server-level security issues such as firewall settings and server-level permissions. The rules also cover many of the requirements for compliance with various regulations.

In fact, documentation of your security status is especially helpful for compliance with some regulations, such as the **PCI Security Standards Council 3.0** standards, that require the reporting of regular database scans. For these purposes, or for your own baselining efforts, the reports can be customised by setting acceptable permission configurations, feature configurations and database settings.



In addition to documentation, the results of the scan include steps to resolve each issue. The system even provides customised scripts which can greatly help you with these steps.

Without a vulnerability assessment tool, you must rely on your own security team being able to keep constantly abreast of all known security issues, and they must be available to review every change to the data warehouse in order to ensure best practices. This is a real burden to any enterprise, and it should be a source of concern: many attacks exploit well-known vulnerabilities that teams have overlooked. Without a tool that can advise on how to remediate issues, you can face the even more maddening case of attacks that exploit issues teams know about but have not yet got around to addressing. The ability of SQL Vulnerability Assessment to generate remediation scripts is a great benefit to database teams under pressure.

Audit logs and log analytics

In addition to the vulnerability assessment, you still need to monitor your data warehouse for suspected abuse and security violations. To track your database activities, **Auditing for Azure SQL Data Warehouse** records events to an audit log. This enables an administrator to be aware of ongoing usage and to analyse and investigate historical activity.

This is a real burden to any enterprise, and it should be a source of concern: many attacks exploit well-known vulnerabilities that teams have overlooked.

Enabling auditing is as simple as selecting a storage account for the logs and turning the feature on. You can set a retention period for the audit logs, which will help compliance with regulations such as the GDPR which requires that a controller "shall maintain a record of processing activities under its responsibility".

Naturally, if you are logging events, you also need a way to review and analyse the log data. Administrators often use simple tools like SQL Server Management Studio (SSMS) or Azure Storage Explorer, or – for the hardcore SQL administrator – just calling the fn_get_audit_file function. These are useful if you only need to check something, but for analysis of large amounts of log data, you will need a tool with more insightful capabilities.

For customers who want more advanced log analytics, or to combine their SQL logs with logs generated from other Azure resources or even on-premises resources, you should consider Azure Monitor. **Azure SQL Data Warehouse** has native support for saving SQL audit logs directly into the log analytics component of **Azure Monitor**. Once the data is in log analytics, you can perform advanced analytics using the KQL query language.

SQL Threat Detection

When you must detect anomalous and threatening database activities as they occur, **SQL Threat Detection** is a powerful tool. This service uses machine learning and behavioural analytics techniques to continuously profile and monitor the behaviour of applications. Any suspicious activity that could indicate a possible malicious intent to access, breach or exploit data in the database will be immediately notified to the database administrator. View threat detection alerts in **Azure Security Center**. They provide details of the suspicious activity and recommend actions on how to investigate and mitigate the threat.

The machine learning methodologies of **SQL Threat Detection** mean that little configuration is required to protect your system. Once you have enabled Auditing (see above), just specify an email address to receive notifications and turn the feature on.

Principle 4: Network isolation is critical

For many years, firewalls have been a key component of network security, especially the so-called three-legged firewall connected to the internet, the DMZ and the intranet. Firewalls are still useful, but increasingly administrators have faced the unpleasant fact that the many exceptions – for VPNs, wireless networks or management ports – threaten to undermine the firewall's effectiveness.

Nevertheless, security boundaries are still important, but you may find it most useful to establish them inside your system. If your network has been breached and the intruder is already inside, isolated security zones help to prevent a single intrusion from compromising your whole network.

Practices

When considering how best to protect valuable resources within your network, it can be helpful to indulge your nightmare: that the system has already been compromised. Hopefully, only in theory, but as we have seen from the Verizon Data Breach Investigations Report, this is all too often true.

You need excellent protection from intrusion but you must also consider how to isolate resources, so a successful intrusion remains limited in scope. Think back to our example of doors and windows – if the intruder has come through the door already, then thank goodness you have your private information secured in a document safe.

There is one key feature in **Azure SQL Data Warehouse** that is invaluable here...

You need excellent protection from intrusion but you must also consider how to isolate resources, so a successful intrusion remains limited in scope.

VNet service endpoints

Often you need to enable Azure resources to communicate securely with each other, with on-premises networks or the internet. In these scenarios, use an **Azure Virtual Network – VNet**. If you need to improve routing efficiency, or to arrange network management in hierarchies, you can subnet the VNet too. A virtual network is scoped to a single region. However, you may also connect together multiple virtual networks from different regions using **Virtual Network Peering**.

In many scenarios, connectivity to your data warehouse will only be appropriate for specific divisions or even teams within the organisation. As I have said earlier, data for analytic systems is often cleansed, enhanced and denormalised in ways which make it especially valuable. **Azure SQL Data Warehouse VNet Service Endpoints** enable the administrator to isolate connectivity to the data warehouse to only a given subnet or set of subnets within your VNets. Not only that but traffic will remain on the Azure backbone network.

In other words, even if your troubled dreams come true and an intruder has compromised your network, they will not be able to access the data warehouse unless they are within an approved subnet.

It is also worth noting that you can separate the roles which provision these **VNet Service Endpoints**. You may configure either the Network Administrator, the Database Administrator or a division of the roles between them.

Principle 5: IDs and permissions need integrated management

There is a high potential for frustration as users juggle numerous logins and passwords across multiple applications and platforms. The truth is, the more IDs we, as users, have to manage, the more we fall back on bad practices such as reusing passwords and making those passwords easy to remember rather than secure. Then, having created these poor-quality passwords, users are reluctant to change them. That's a sure recipe for a compromised system. But forcing complex password management on users is an equally certain path to discontent.

Your aim should be to connect all your users with all your apps and data seamlessly, whether they are logging into **Azure SQL Data Warehouse**, **Office 365**, Dropbox, Adobe Creative Cloud, Salesforce or other SaaS applications.

Increasingly, network and data warehouse administrators not only need to define permissions by a user's identity but by location, device and application context. A user who has access to customer data from a CRM tool may not have access from a data analysis platform. Or a user with access to financial data while in the United States often will not work with the same permissions when travelling in Asia.

Practices

Azure Active Directory

Azure Active Directory (Azure AD) is the Microsoft cloud-based identity and access management service. Azure AD helps your employees sign in and access external and internal resources.

With **Azure AD** authentication, you centrally manage the identities of Microsoft service, including **Azure SQL Data Warehouse**, users in a central location. This not only simplifies permission management, it also helps to prevent the proliferation of identities that I described earlier and enables the rotation of passwords in a single place.

In fact, Azure AD can eliminate storing passwords by enabling integrated **Windows Authentication** for users and token-based authentication for applications connecting to **Azure SQL Data Warehouse**.

Not only do you have central permission management for individual users and groups, you may also set permissions for external or guest users with and without **Azure AD** access.

Principle 6: Sensitive data needs fine-grained access controls

There are numerous aspects to securing access to resources in your data warehouse, but what about the nature of the *raw material* – what about the data?

Data is the lifeblood of modern businesses, flowing through multiple systems. When this complex, varied data is integrated and structured for comprehensive analysis, the data warehouse is at the heart of that process. However, not all data is equal, especially when it comes to compliance with data privacy regulations which aim to protect a customer's personal information.

It is important, therefore, to locate systems which store sensitive data and identify which data is personal. Within a data warehouse only a subset of tables will contain personal data and, even within those tables, only specific columns or rows.

In addition to identifying and labelling sensitive data, we have to ensure that access is carefully controlled. After all, there are legitimate uses of personal information. You want to allow just the right users to see just the right data.

For example, in a hospital, medical staff should only be allowed to see patient data that is relevant to their own patients, not every patient's data. Research staff may see data from all patients but often have no need to see personally identifiable information for their research.

When I write about sensitive data, you may think first of social security numbers or national identifiers, or even credit card numbers. Well, this data is certainly sensitive, but in my opinion it should not even be stored in the analytic data warehouse because these data items have little analytic value. They are, by design, unique to each user and are therefore modelled effectively by your surrogate keys. Yes, there is information embedded in a credit card number about the card type and issuing body, but this can be more usefully extracted and codified during the ETL process.

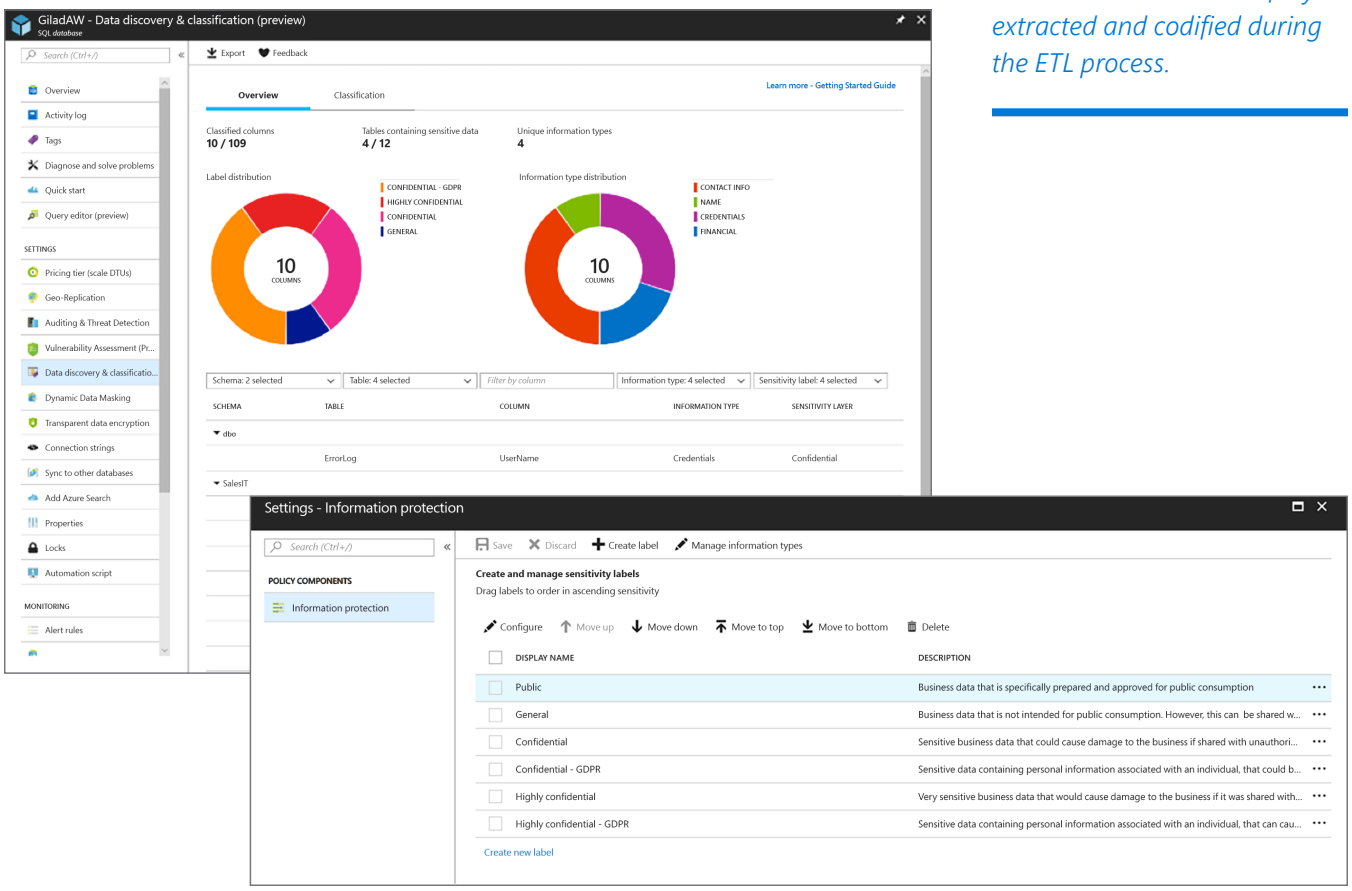
Practices

Identifying and classifying sensitive data

Traditionally, identifying sensitive data has involved data administrators or data stewards poring over E-R diagrams or querying metadata (using sys.columns) for column names which indicate personal data, such as Name, Birthdate, and so on. These sensitive columns would be tagged using extended properties in the database and typically added to an externally-maintained data map for future reference.

This is often a difficult and error-prone process. Not all columns of sensitive data are so easily identified by name and some, such as Compensation, Evaluation or Level in a Human Resources system, may not be obviously sensitive to a non-specialist.

SQL Data Discovery and Classification is a tool for discovering, classifying, labelling and reporting the sensitive data in your databases. It is built into SQL Server Management Studio (SSMS) and also available on the Microsoft Azure Portal.



The classification engine scans your data warehouse, identifying columns containing potentially sensitive data. You should review the list of recommended column classifications and accept or change the recommendations. The engine will also attempt to identify the information type, such as *Contact Info* or *Credentials* and change these as needed too. Finally, as a best practice, tag columns with a Sensitivity Label such as *Confidential* or *GDPR*. In this way, you have a useful record of why the data is considered sensitive in a given scenario.

Administrators can query for sensitive information using the extended properties: `sys_information_type_name` and `sys_sensitivity_label_name`. But of course, the system also provides detailed reports for compliance and auditing.

Without a classification engine like this, you must again rely on an experienced internal team to know about requirements in detail and to review your architecture and all changes in order to govern your system.

Object-level permissions

It is important when deciding on the permissions for any account to consider the principle of least privilege. Users and applications have the minimum privileges required to perform their task. Don't be blinkered by the assumption that all users have the same needs.

For example, an administrative assistant (often a junior or contract worker) may need to view information from the employees table in order to be able to contact staff when needed. The relevant permissions can be granted by the data warehouse DBA as follows:

```
GRANT SELECT ON Employees(EmployeeID, FirstName, LastName,
Phone, Email) TO AdminAssist;
```

I'm using code here to make the example clear, and because DBAs do like to run or schedule scripts when setting properties on many objects, but this process could also be completed in the user interface of **SQL Server Management Studio**.

The administrative assistant has been granted permission only to see certain columns. If they attempted to select all columns, they would see the following error:

```
SELECT * FROM Employees;

Msg 230, Level 14, State 1, Line 12

The SELECT permission was denied on the column 'SSN' of the
object 'Employees', database 'HR_DW', schema 'dbo'.
```

While administrative staff are unlikely to write in SQL, the script is shown here because the same code might be generated by a selection made in a BI tool or a connected spreadsheet.

Column-level permissions, such as these, are necessary for any reasonable data warehouse privacy protection. However, column-level permissions are not sufficient for good data warehouse privacy. You need row-level security (RLS) too.

RLS controls access to rows in a database table based on the characteristics of the user executing a query. In this way, only database users that have an identified need to access the content of a specific database row will be granted that access. For example, workers need to access only those data rows that are pertinent to their department, or a nurse's access might be restricted to relevant data only for her assigned patients. Row-level security greatly simplifies the design and coding of this type of access management and security within the application.

Elevating data access should be a career death-wish for administrators in the 21st century. But I'm still amazed how often I find DBAs elevating privileges for demanding users because it's easier than configuring and debugging all the relevant object permissions correctly. Don't do it, even for fear of offending the demands of the business. Choose a platform which offers both an excellent toolset for setting permissions – like SQL Server Management Studio – and services such as SQL Vulnerability Assessment which can intelligently assist you in identifying badly configured rights.

Without native RLS, you typically need to define access restrictions in the application tier of your architecture, but remember, that means you must define it in every application. In the world of self-service, where users may well be 'bringing their own' applications, that is almost impossible to ensure. Another approach, without RLS, is to define database views that restrict access. But then a view needs to be built for each scenario, and you must now manage permissions on the view too. The approach is cumbersome and error-prone.

With native RLS, the access restriction logic resides simply in the database tier. The restrictions are applied every time that data access is attempted from any tier. This makes the security system more reliable and robust by reducing the system's surface area.

Principle 7: Business continuity is a security and privacy issue

It's one thing to protect your data when your systems are running smoothly, but you also have to consider its resiliency and availability in the event of an adverse incident. For customers and users, it's infuriating enough when a system is down. They want service restored quickly and not just to get back to work. In the darkness of downtime, they cannot assure themselves that their interests have not been compromised: the wait is too often a protracted agony, relieved only when the application is back up and running.

The GDPR, focused as ever on the concerns of consumers, explicitly refers to this critical need by requiring that the organisation "implement appropriate technical and organisational measures" that include "the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident".

Practices

As you would expect, the **Azure SQL Data Warehouse** service automatically performs regular database backups. The service guarantees **Point-in-Time Restore** from these backups for a certain scope of recovery.

However, in many cases you could need longer term storage. For example, in considering the GDPR, the Belgian Data Protection Authority has suggested that records of processing activities be kept for five years. Long-term retention for backups is available by storing **Azure SQL Data Warehouse** backups in an **Azure Recovery Services** vault for up to ten years.

Azure SQL Database also offers an **Active Geo-Replication** feature, which provides a database-level recovery solution with low recovery time. **Active Geo-Replication** enables the configuration of up to four readable secondary databases in the same or different regions. But, it is also important to note that in all its business continuity solutions, Azure respects data residency requirements. While Microsoft may replicate data to other regions for data resiliency, it will not replicate or move customer data outside the defined geo-boundary. Customers, however, can move, copy or access their customer data from any location globally.

Conclusion

In some ways we have come full circle with these principles. The first two – that security and privacy are separate but related practices and must be built on the best platform – apply in many IT scenarios. So too with our last proposition – that resiliency is a governance issue, too. These fundamentals, along with your users' need for simple but robust authentication, should influence your choice of cloud provider.

Those principles specifically related to data warehousing needs would also demand a sound choice of cloud provider. Moreover, they require you to choose a database platform that supports security and privacy as a foundational component of its design.

Column- and row-based security are not optional for a well-governed data strategy. In an age when IT departments are increasingly supporting data science pipelines, it is not too much to ask that machine learning should support your own work too. Look for intelligent discovery and tagging of sensitive data, along with continuous monitoring for anomalies.

The **Azure SQL Data Warehouse** provides all these capabilities, while in many cases other cloud providers and data warehouse platforms fall short. Perhaps even more important for strained IT and data governance teams, these capabilities are well integrated into a coherent management environment.

There are naturally other considerations when choosing a cloud data warehouse, ranging from modelling, to support, to cost of ownership. Nevertheless, your security and privacy strategy is a fundamental requirement that cannot afford to be compromised.