International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Data Location and Data Type Considerations for Identifying Security Measures

Anand Athavale

Independent Researcher, Decades of Industry experience in Data Management
andyathavale@gmail.com

Abstract
Defining security measures is a complex process and it involves more than just InfoSec teams. Effective security measures help reduce probability and impact of a ransomware or a compliance breach. In the real world, many of those security measures are common, but those are scattered among different IT practitioners. This article elaborates on data location and data type considerations to help define a subset of security measures which cut across storage, application, and security administrators. Data location type, while driven by the purpose, may not be specified by the data owners and data operators but instead left to the discretion of IT staff, to fit within scale, budget, and architectural considerations. Data type is also derived by the data purpose but sometimes selected by the data owners or creators, based on the choices available to meet the purpose. IT and security practitioners getting ready to define security measures for data in their organization will benefit from this article, by learning the role data location type and data type plays in identifying security measures.

Keywords: Information Security, Data Compliance, data security measures, Ransomware resilience

Introduction

As the data owners or creators create any data item, the data location is already established, a good percentage of times. Unless it is a completely new application, or a project requiring design from the ground up, there are pre-communicated expectations on where a specific data item may be residing.
A document such as this, has a designated location pre-identified and could be on-premise file server, on-premise SharePoint or similar system, or, a cloud Software-as-a-Service based application like OneDrive, Google drive or similar.
If it were a new email message, it would have its own application already set up. Of course, the data location as seen by the creator may be totally different than how it is being interacted with by storage administrators. In the case of an email, the email writing person would interact only with the Outlook or similar application. The storage administrators however would be interacting with MS Exchange server deployment and the storage used for the same.
More purposeful business applications would have already deployed necessary databases and applications to interact with those, stored on On-premise storage, or, in the cloud. Modern applications may be using platform-as-a-Service applications like AWS RDS or Azure Cosmos DB.

Data Type

Security measures can be looked at from the data type perspective as follows.

Unstructured Data

This is the most open and undefined data type. Data stored on any laptops, file servers and similar locations can be qualified as unstructured data. While there may be some hierarchy to the organization of the data, it is very loosely defined and controlled. The interaction of the data may be restricted depending on the type of applications, but most of text-based types could very well be read and manipulated using simple and general operating system level tools and commands. The example of application specific file formats is MS word, power point, excel and PDF files. There could be controls within the application to define data security posture, but those are very minimal. Example of such a method is Excel allowing authors to protect certain tabs with passwords.

Structured Data

Structured data are typically databases or record management systems. There are many types and kinds of databases and similar applications or systems. While most of them get stored on some type of hardware storage, there are a few variations, which are stored in memory. Structured data types mostly have a designated way or tools for data definition, referred to as Data Definition Language (DDL) and data change or manipulation, referred to as Data Manipulation Language (DML) [1]. Security measures however need to be considered on the storage side and in addition, within the application which interacts with the structured data. Considering only one or the other is not sufficient.

Semi-Structured Data

At a granular level, semi-structured data type is not that different from unstructured data type. The same files or file types considered as unstructured data can reside in a semi-structured data source. As an example, this very file can reside on a laptop, inside a folder on a file share, on One drive, or a SharePoint folder even. The application that stores it though may be considered semi-structured. The systems or applications which allow for better control of the unstructured data at a granular level could be considered semi-structured. So, this file on anyone's laptop may be considered as unstructured data. But, the same file, organized and stored on say Google drive, with properly set controls for access could be considered semi-structured. The same file can be accessed or changed using different flavors of the same applications. But, the security measures are controlled by the location, not the flavor of the application.

Data Location Type

Data Location has many facets which determine the process of data security measures. Again, to establish the scope of the same, it is important to note that there are many connection types involved when it comes to Data location. Those include direct attached storage (DAS), network-based storage (NAS) and storage access network (SAN). While these also play a role in overall security measures, these are closer to infrastructure security and configuration management than specific data item security measures. These are not in the scope of this article.

For data locations at a very basic level, the various aspects could be grouped as follows.

Storage layers

Typically, lower layers of storage are less easily accessible than the upper layers. Due to this, higher level storage layers require additional considerations for data security measures.

Layer 1 - Disk, Tape, Other devices

Disks or Hard Disk Drives are usually part of servers or what are referred to as storage systems like arrays.

When defining data security measures, following elements should be considered.
Theft protection

- Locking mechanisms for the detachable disks Self-encryption
- Physical inventory and Tracking mechanisms
  Destruction protection
- Mechanisms to prevent erasure
- Protection from overwrites using low level tools

Tapes

Tapes typically get used for copies of data and not for on-demand interaction with the data. The considerations mostly cover physical access as tapes are by design separated from the original storage systems. However, tapes still require security measures [2]. Some of those are listed as an example below.
Theft protection

- Tape Device encryption
- Host based encryption Destruction protection
- Tape level prevention settings for overwrite
- Backup application-level prevention settings for overwrite

Other devices

Other devices such as USB drives and CD ROMs are not scalable and get used less in organizations. However, if used, physical security becomes most important here. There are devices with digital and/or physical key locks for security which should be considered depending on the criticality of the data when identifying data security posture processes. Self-encryption becomes also necessary. There is also a side-angle equivalent to what was referred to as "dumpster diving". Physical discarding of such devices becomes another item for consideration given that fallen into wrong hands, it could create security risk for theft.

Layer 2 - Volumes, Logical Drives, WebApps, Databases, Subscriptions

Volumes

For manageability, storage management applications have a concept of volumes. From this layer, these elements are at a more logical level than a physical and are more software controlled than hardware controlled. Operating systems typically provide volume management by default. However, there are other volume management options available too. Connecting this layer to the hardware level disks also has various options. Multiple volumes can be "carved" out of a single disk. A single volume can also span multiple disks for various reasons like overcoming a single disk size limit, protecting against disk corruption and increasing performance for the input/output or I/O. At a very high level, both servers and storage systems have the concept of a volume.

For Cloud based storage, while there is a concept of volumes within server instances, it has some variations. Internally, cloud service providers may have another layer or a different layer, but the top most layer exposed to the user side storage admins is a subscription. While subscription does not carry the same concept of a volume, it does sit at a level which allows for bulk level access or deletion. There are other layers such as resource groups and storage accounts, which could also be clubbed to be at the volume level.

Databases typically may not have a matching concept either as compared to storage layers like disks and volumes. Some database technologies however do allow for volume management which directly correlate

to database storage concepts like data files. However, these are optional. Using standard storage concepts at this level and relying on next layers like file systems for correlating concepts like data files depends on the architecture and storage chosen for databases. A layer like file system gives more flexibility but it also makes the data items easily accessible and increases the number of tools for interaction. As an example, commands like dir for windows and ls for Linux are available to traverse and find data, which is not possible easily at the volume level. This in turn increases the number of data security measures to be considered reiterating earlier point that as you go upwards in the storage layer, accessibility increases but so does the number of elements to consider for data security posture.

There are many more applications and constructs comparable to volumes. For example, SharePoint has a concept of a Web Application. It typically has both, the application code elements interacting with the data items in some way and the data items itself. However, for accessibility, the next layer is required which is typically a site collection with one or more sites. From this layer onwards, following are some typical measures which need to be considered [3].

Theft Protection

- Encryption
- Limiting tools access (ex. To commands like disk dump or dd)
- Limiting Identity access
- Requiring additional controls for destructive operations

Destruction Protection

- Limiting tools access (ex. To commands like disk dump or dd)
- Limiting Identity access
- Requiring additional controls for destructive operations

Layer 3 - File Systems, Shares, Site collections/sites, Accounts, mailboxes

The next level is often the final level where storage administrators have any type of control for defining or ensuring data security posture. For upper layers than this, the onus of data security posture typically falls on application administrators and the data owners or creators, depending on data type.

File systems are typically the next level constructs on top of volumes. File systems enable the method of access to the individual or collection of data items along with organization and structure of those. Irrespective of whether it is a server-based file system, or part of a storage system like arrays, they typically have a security style defined. These styles are tied to the identity systems which define the “users” of the data items. It is important to note that from here on, there are many possibilities of what a file system looks like and how a file system maps to a share. In some cases, the security styles may differ in those two constructs.

File System and/or Share Security Styles

- NTFS
- Unix
- Mixed

There are many more constructs at similar level, but significantly different from file systems. In the cloud object storage, there is a concept of a bucket or a blob. In SharePoint site collection with a single or multiple sites is somewhat comparable to a share or a file system. However, the security style is completely distinct and more granular. For email systems, there are mailboxes. Some mailboxes are individual and some are shared. There is also a concept of public folder mailboxes in Exchange as an example [4]. Again,

when to use which type is dictated by the purpose of the mailboxes. However, each of the types does have an impact on establishing data security posture. In this paper, we are not discussing permissions in depth. That will be covered when we discuss data exposure.

Accessibility options

Defined vs. open/undefined access

Depending on the location type, storage type and storage layer, there are defined methods to access and/or alter data. Here is an example.

Let's consider a database, for instance, Oracle. Now, Oracle has a construct of data files, which could be physical files, if those are stored on a file system. Now, even if it is an Oracle database file, there is no prevention from treating those as another file system file. Of course, there are mechanisms in Oracle, the application, to track or flag any "out of the band" changes. But those may not be present for all database applications. In such a scenario, the file system storage layer considerations we discussed in Layer 3 become applicable. Oracle also allows "raw devices" or volumes as data files as part of ASM (Automatic Storage Management) [6]. In that case, considerations of layer 3 may not apply. This is the reason, accessibility method which is tied closely to data type needs consideration for data security measures. In the case of Oracle, some security administrators may lean towards ASM vs. file system as ASM has lesser security measure considerations. On the other hand, storage administrators may find it easier to manage and hence may lean towards a file system for Oracle.

Tailored Application Access and Data Type

Oracle is a good example, also for tailored application access. Typically, besides some internal operations, Oracle defines a application specific interaction method called SQL, which could be segregated into create, select (query), update (modify) and delete type of operations. Hence, the application itself gives a segregation into read and other operations, which in turn, gives ways and means to define data security measures within the access method itself.

However, some applications have mixed methods. Take an example of a SharePoint stored file. That file can be modified and viewed using traditional Microsoft applications like MS word etc. However, sitting in SharePoint, there are additional operations possible on that file like Check-in and Check-out [6]. Due to such variations, the data location type matters beyond just the data type during data security measure identification process.

Conclusion

Looking at data types and data location types, it becomes clear that data security measures can not be established in isolation by any one team, or, just at one level or the other. Whether after-the-fact, or, proactive, data security measure identification process needs a holistic consideration. As seen from the examples and layers, leaving out any one accessibility method, storage layer, or data type could leave gaps in the information security. Hence, involving all teams and establishing a data security measure identification process at the organization level becomes a must-have and not a nice-to-have. New concepts like Infrastructure as a code and containerization of applications only adds to the necessity of a holistic consideration.

It is important to acknowledge that while we discussed the data type and data location type, what that data holds or contains, has a significant amount of impact on the data security measures identification process.

References

1. Aniket Thakur, Difference between DML and DQL statements in SQL (2014), https://opensourceforgeeks.blogspot.com/2014/10/difference-between-dml-and-ddl.html (June, 2019)
2. [Product Documentation, No author], DFSMSdss Storage Administration Guide (Securing your tape backups), https://www.ibm.com/docs/en/zos/2.1.0?topic=dfsmsdss-securing-your-tape-backups (June, 2019)
3. Arnold Johnson, Kelley Dempsey, Ron Ross, Sarbari Gupta, Dennis Bailey, Guide for Security-Focused Configuration Management of Information Systems (NIST Special Publication 800-128), (August 2011, updated October 2019) https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-128.pdf, (November 2019)
4. Paul Cunningham, Office 365 Groups vs. Shared Mailboxes, Practical 365 by Quest, (November 2017), https://practical365.com/office-365-groups-vs-shared-mailboxes/ (May 2019)
5. Oracle Support Forum User], ASM vs OCSF2, Database Software, (Jan 2016), https://forums.oracle.com/ords/apexds/post/asm-vs-ocfs2-6657 (May 2019)
6. [Admin], SharePoint 2010 Permissions management Guide, LighteningTools (August 2012), https://lightningtools.com/permissions/sharepoint-2010-permissions-management-guide/ (May 2019)