DRAFT TLP:AMBER:CTID

Analytics Blueprint Best Practices and Lessons Learned for Developing Cloud Analytics

1	Intro	duction		1
2	Ana	ytic Devel	opment	2
	2.1			
	2.1.	. Initia	al Approach	2
	2.1.		eted Approach	
	2.2			
	2.2.		elation Analytics	
	2.2.	2 Cross	s-Cloud Analytics	5
3	Case			
	3.1		g a TTP	
	3.2		/ Emulation	
	3.3		nalytic and Map to TTP	
	3.4		Analytic	
	3.4.		a Analytic Components	
4	Infra	structure	Setup	12
	4.1	Infrastruc	ture Requirements	13
	4.2	Architectu	ure	13
	4.3	Deployme	ent	13
	4.3.	Clou	d Service Rules of Engagement	14
5	Itera	tive Appro	oach	14
	5.1	Exercise A	Adversary Behavior	14
6	Sum	mary		15
	6.1	Project O	utcomes	Error! Bookmark not defined
	6.2	Challenge	25	Error! Bookmark not defined

DRAFT TLP:AMBER:CTID

6.3	Cost Considerations	Error! Bookmark not defined.	
6.4	Sharing Cloud Analytics	Error! Bookmark not defined.	
6.5	Future Opportunities & Research	Error! Bookmark not defined.	
7 Ap	pendix A – Terminology	15	
8 Ap	ppendix B – Cloud Offensive Toolkits	18	
8.1	Azure	18	
8.2	AWS	18	
8.3	GCP	18	

1 Introduction

There is a significant shortage of cybersecurity talent worldwide, with estimates around 500,000 unfilled positions in the US, and roughly 2.7 million unfilled cybersecurity jobs globally¹. Existing defenders are stretched thin and struggle to achieve desired visibility of adversary behavior in cloud environments.

Cyber analytics are a data-centric approach to cybersecurity based on collecting and analyzing large volumes of event data collected from heterogenous sources. High quality analytics can quickly identify instances of malicious behavior in a sea of data, reducing the time to detection and focusing the defenders' attention on the most important activities.

The creators of analytics must carefully balance the impact of false positives, since analytics with low signal-to-noise ratios require more staff time to process and draw attention away from true threats. Projects such as Sigma² provide a structured format for detection analytics that can be translated to platform-specific queries for platforms such as Splunk, CrowdStrike, Google Chronicle, and Microsoft Sentinel, among others³.

The Center for Threat-Informed Defense launched the Cloud Analytics project to advance the state-of-the-art in cyber analytics for cloud platforms. The initial focus is on Azure and Google Cloud Platform, but the results generalize to other cloud platforms. The project leveraged the MITRE ATT&CK® Cloud Matrix¹⁹ as a reference for cloud-specific adversary behavior. The deliverables include a set of cloud analytics for key tactics, techniques, and procedures (TTPs) as well as best practices and lessons learned during the project. This blueprint document provides guidance on the processes, challenges, and workflows that can be used by defenders to create their own cloud analytics. The intended audience is creators of cyber analytics for cloud environments, but analytics users will also benefit from this document.

¹ NICE FactSheet Workforce Demand Final 20211202.pdf (nist.gov)

² SigmaHQ/sigma: Generic Signature Format for SIEM Systems (github.com)

³ <u>SigmaHQ/pySigma</u>: <u>Python library to parse and convert Sigma rules into queries (and whatever</u> else you could imagine) (github.com)

DRAFT **Exercise** Stand up Develop **Analytics** adversary analytic Blueprint test range behavior Best practices and lessons learned for developing Refine cloud analytics reduce false positives

Figure 1. Learn-by-doing Project Approach

This document is grounded in the practical experience of iteratively developing, testing, and refining cyber analytics in real cloud environments.

2 Analytic Development

2.1 Process

Creating a cyber analytic is a complex task that requires balancing false positive risk and detection granularity of the intended target. The Cloud Analytics project experimented with two approaches, an initial, exploratory approach that encountered challenges, and a second, targeted approach which built upon the lessons learned from the first. Ultimately the targeted approach produced better results, but both approaches provide valuable lessons.

2.1.1 Initial Approach

The initial project approach consisted of identifying cloud adversaries, developing emulation plans, executing those emulation plans, and sifting through the results to find meaningful signals to build analytics on.

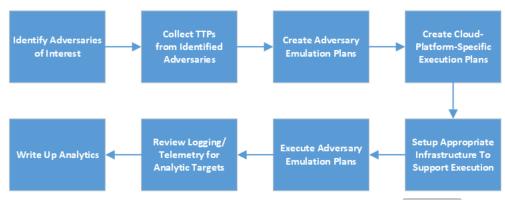


Figure 2. Initial approach to developing cyber analytics.

The initial approach proved challenging and produced disappointing results. The first issue was the limited amount of open-source intelligence for cloud-specific adversaries. The Cloud Analytics project team initially focused on TeamTNT, a threat actor known for cryptocurrency mining campaigns on cloud infrastructure. To produce a well-rounded adversary profile, the project team augmented the TeamTNT adversary emulation plan with additional ATT&CK TTPs for exploiting container-based platforms. The complexity of the environment combined with the breadth of TTPs—especially resource enumeration techniques—generated an overwhelming amount of innocuous log events.

Even under ideal conditions with minimal cloud assets, cloud platforms such as Azure and GCP can produce hundreds or thousands of log events per hour. As a result, the project team was searching for a needle in a haystack to find useful signals.

2.1.2 Targeted Approach

In contrast, the targeted approach started with the identification of a small set of cloud-native TTPs. Compared to the TeamTNT emulation plan, this approach reduced complexity and logging noise. As a bonus, it also simplified the mapping of analytics to ATT&CK techniques, since each analytic was paired with a specific TTP at the start of the process.

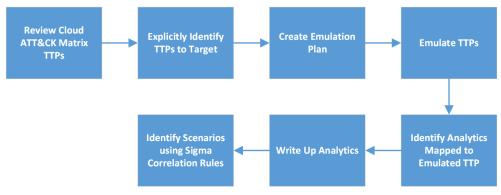


Figure 3. Targeted approach to developing cyber analytics.

2.2 Analytics

Analytics for the project take the form of Sigma rules. Sigma is a common language for defenders to define log events of interest. Conversion engines, such as sigmac or PySigma, translate rules from the Sigma language into the appropriate language for a specific platform such as Google Chronicle, Microsoft Sentinel, Splunk, or Elastic Search. This standard is powerful because it enabled the Cloud Analytics team to create one set of vendor-agnostic analytics that can be deployed in a variety of environments.



Figure 4. Sigma rule conversion workflow²¹

The project team created detection rules designed to detect specific events mapping to specific ATT&CK techniques. Individual detection rules are prone to false positives and should be considered building blocks of a layered detection program. In 2022, the Sigma project released a new feature for creating correlation rules—rules that match multiple, related events—due to timing, a small set of preliminary correlation rules were integrated into the Cloud Analytics project. This is an area of future research and the applications of correlation rules are discussed in the next section.

2.2.1 Correlation Analytics

Sigma Correlations¹ is a new addition to the Sigma standard that builds on atomic rules for "the expression of aggregations and relationships between rules." The correlation concept allows for expressing temporal relationships between detection rules and grouping detections by attributes such as hostname or username.

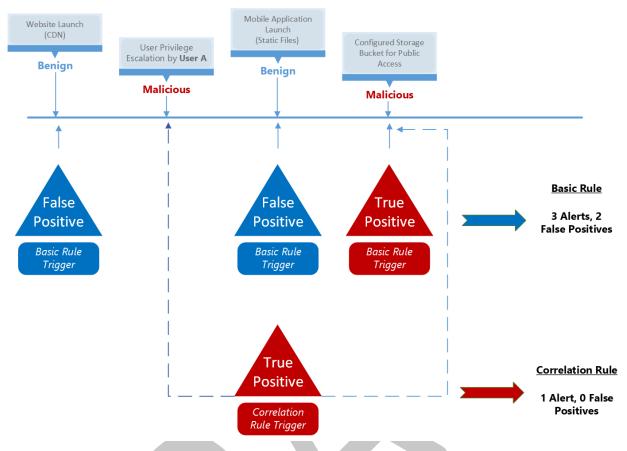


Figure 5. Diagram comparing standard detection rules vs correlation rules.

For example, Figure 5. Diagram comparing standard detection rules vs correlation rules. Figure 5 shows a basic analytic that detects when a cloud storage container is modified to allow public access. This analytic may generates false positives if, for example, a new web site is launched and its storage bucket is switched from private to public as part of the release. The correlation rule depicted above detects the specific sequence of 1) a guest user being granted administrator access, and then 2) that same guest user modifying a storage container to allow public access. The correlation rule provides a stronger signal of potential malicious activity.

2.2.2 Cross-Cloud Analytics

Cloud platforms (such as Microsoft Azure and Google Cloud Platform) have significant differences in terms of organizational structure and resource organization. However, there are high-level cloud concepts that can usually be mapped to the specifics of each platform. Theoretically, Sigma is a common language that maps to these cross-cloud concepts and can be converted into platform-specific rules. In practice, there are significant limits at play when creating cloud-agnostic Sigma rules. The concepts are generalizable (creation of additional resources, autoscaling usage, modification of access controls, and cross-cloud service references are immensely helpful in identifying equivalent services²), but service names, logging

architectures, and detection queries all vary significantly between platforms. These variations require the creation of distinct rules for each cloud platform. Cross-cloud service references are immensely helpful for mapping out equivalent services across cloud platforms³.

Other projects, such as Elastic ECS, are working to provide a common schema to map log events across platforms. For example, an ECS object will store information about a user under the "User" field set regardless of the source. Azure may provide the user metadata via the User Principal Name, Google Cloud may provide the user email under a principalEmail field, however the ECS processor for both log sources would map the data to the appropriate ECS attributes. The common schema allows for analysis across all data sets.

3 Case Study

To further examine the analytic development process, we will work through an example using the targeted approach discussed earlier.

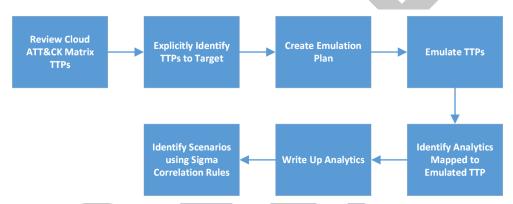


Figure 6. Targeted approach to cyber analytic development.

3.1 Identifying a TTP

Review the appropriate ATT&CK Matrix for your area of interest. As previously mentioned, this project used ATT&CK Cloud Matrix ⁴.

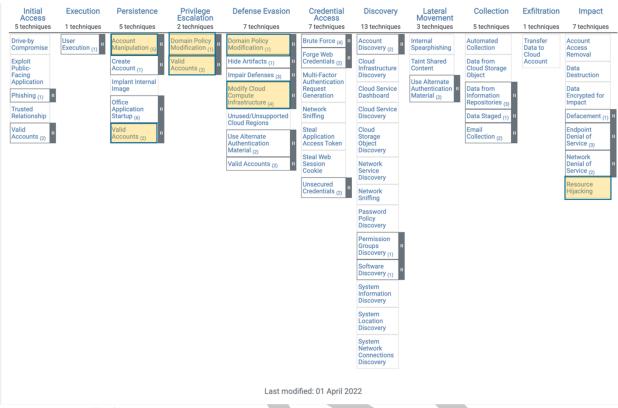


Figure 7. ATT&CK Cloud Matrix

For this example, we select T1578: Modify Cloud Compute Infrastructure. While the T1578 TTP covers a wide range of activities, we will focus on modification of storage buckets to allow public access. The full ATT&CK entry for T1578 can be found in Figure 8. ATT&CK TTP T1578.

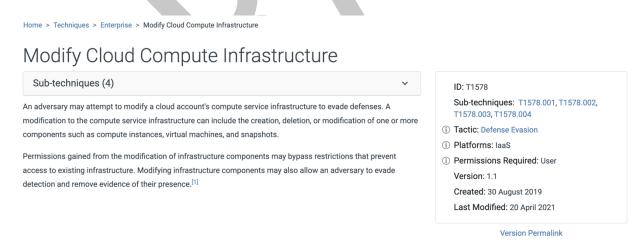


Figure 8. ATT&CK TTP T1578

3.2 Adversary Emulation

Now that we have identified a specific TTP and resource target, the next step is adversary emulation. Both Azure and GCP provide documentation on using their CLI tools to make a storage container public.

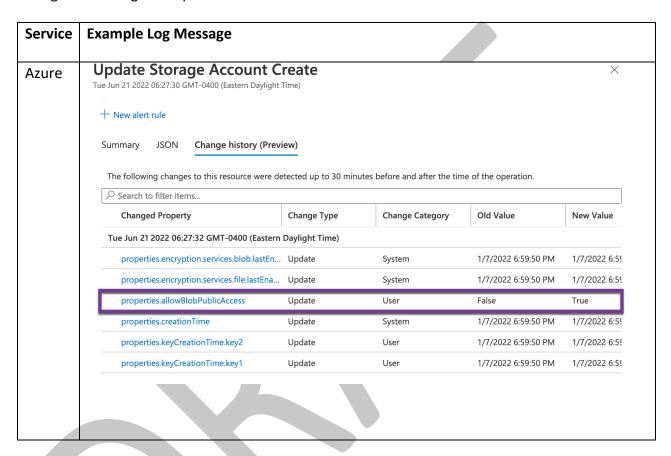
The respective commands for Azure and GCP are below.

Service	Command
Azure ⁵	az storage account update \
	name <storage-account> \</storage-account>
	resource-group <resource-< td=""></resource-<>
	group> \
	allow-blob-public-access
	true
GCP ⁶	gsutil iam ch
	allUsers:objectViewer
	gs://BUCKET_NAME

Figure 9. Commands to make a storage bucket public on Azure and GCP, respectively.

3.3 Identify Analytic and Map to TTP

Once the emulated attack has executed, review the appropriate documentation and logs to identify any significant events that resulted from it. The following table provides examples of a storage blob being set to public for Azure and GCP.



```
GCP
                "protoPayload": {
                  "@type": "type.googleapis.com/google.cloud.audit.AuditLog",
                  "status": {},
                 "authenticationInfo": { --
                 "requestMetadata": { --
                  "serviceName": "storage.googleapis.com",
                  "methodName": "storage.setIamPermissions",
                  "authorizationInfo": [--
                  "resourceName":
                  "serviceData": {
                    "@type": "type.googleapis.com/google.iam.v1.logging.AuditData",
                    "policyDelta": {
                      "bindingDeltas": [
                          "action": "ADD",
                          "role": "roles/storage.objectViewer",
                          "member": "allUsers"
                 },
                 "resourceLocation": {--
```

Figure 10. Examples of storage blob becoming public for Azure and GCP.

For the mapping component, as the targeted approach starts with an ATT&CK TTP, the ATT&CK mapping for the TTP is already complete. In this case, the analytic will be mapped to T1578.

3.4 Write Up Analytic

The write up component involves compiling the information gathered in previous steps into a Sigma rule. The Sigma project provides useful documentation, as well as third-party blogs and books⁷. Sigma is YAML-based and provides a schema. For creating/editing Sigma rules, the Visual Studio Code extension sigma⁸ is useful in autocompleting tags, providing code snippets, auto generating UUIDs, and performing preliminary static analysis.



3.4.1 Sigma Analytic Components

The Sigma project documentation²¹ is a valuable reference, but it is worth calling out a couple parameters from rules whose applicability may not be clear for cloud environments.

Attribute	Description	Azure Value	GCP Value
logsource	Service dependent log service name.	<pre>{product: azure, service: azureactivity}</pre>	<pre>{product: gcp, service: gcp.audit}</pre>

Sigma recommends reviewing existing logsource values rules for cloud platforms logsource parameter for cloud services. Reviewing the existing Sigma -> Platform conversion tools is helpful to see what current attributes are supported.

4 Infrastructure Setup

A properly configured cloud infrastructure environment is the first requirement for developing cloud analytics. If an environment is not readily available, setting up infrastructure can be a

non-trivial step requiring business funding and approvals to provision the required resources. It is important that the provisioned environment is not overly restrictive—in permission access, feature availability, and metric and logging collection—to serve the desired purpose of cyber analytics research and development. The Cloud Analytics project experienced these issues first-hand: the novelty of the project and cybersecurity concerns created bureaucratic hurdles and delays before the project could begin in earnest. The solution involved setting up an isolated set of cloud environments, which is discussed in detail below.

4.1 Infrastructure Requirements

Attack emulations directly against cloud infrastructure is a new concept to most organizations. To minimize bureaucratic overhead, the project team identified the following requirements for a cloud test range.

- Full administrator permissions over the cloud account(s).
- Separate and standalone cloud service account(s), with no connection to production corporate environments.
- Sufficient budget for cloud service usage.

4.2 Architecture

The required infrastructure components may vary depending on the specific cloud platform used, as well as the goals of the project. For the Cloud Analytics project, the following components were used for the base deployment. Note that some adversary emulation steps may create additional infrastructure. Both Microsoft Azure and Google Cloud Platform (GCP) services were used during the project for perspective on cross-cloud analytic development.

- Windows Virtual Machine (VM)
- Linux VM
- Container-based application deployment (Azure Container Instances, GCP Compute)
- Secret Manager (Azure Vault, GCP Secret Manager)
- Storage Service (Azure Blob Storage, GCP Storage)
- User Management (Azure Active Directory, GCP IAM)
- Virtual Networks resource communication
- Auditing infrastructure for telemetry storage
- Elastic Cloud cluster for log ingestion and analysis

4.3 Deployment

When possible, automated Infrastructure as Code (IaC) tools, primarily Terraform, were used to define, provision, deploy, and maintain infrastructure. Terraform uses a declarative syntax for specifying cloud resources such virtual machines and virtual network. Terraform makes deployment of cloud infrastructure simple and repeatable, immensely reducing the labor required for managing cloud infrastructure.

4.3.1 Cloud Service Rules of Engagement

Cloud services typically have rules of engagement for performing emulated attacks and pen testing against cloud infrastructure. References to policies for different cloud providers are shown in **Table 1**.

Cloud ProviderReferencesAmazon Web Services (AWS)9Google Cloud Platform (GCP)10, 11Microsoft Cloud (including Microsoft Azure)12

Table 1. Cloud Provider Rules of Engagement

An incorrectly provisioned account may result in one or more of the following issues:

- Inability to create/modify user accounts during attack emulation for different scenarios.
- Inability to view all logged account activity.
 - Most corporate cloud configurations limit visibility of logging and metrics to resources under your control.
 - As a result, you will have limited visibility to logged events logged by protected resources.
 - For example, attempts to compromise secret storage services, such as Azure KeyVault, will not be viewable if the user viewing the logs does not have access to the KeyVault resource.
- Triggering internal security processes that identify malicious activity.
 - o If the organization perceives the adversary emulation as a real threat, this will likely result in disabling/removal of related accounts and infrastructure.

5 Iterative Approach

5.1 Exercise Adversary Behavior

Adversary emulation is a critical component when generating metrics to create and validate cyber analytics. The adversary emulation plans used in the project follow the common schema defined by the Center's Adversary Emulation Library project¹³ for interoperability with existing tools such as CALDERA¹⁴.

To develop adversary emulation plans (AEPs), the project team initially consulted with MITRE subject matter experts and reviewed existing adversary emulation plans, as well as research related to adversary targeting of cloud environments. The team discovered limited cloud-specific information, as most known adversaries typically attack hosted resources, such as a Windows Server instance, regardless of whether it is hosted in the cloud or on-premises. Given the project timeframe and focus on detection, the Assume Breach Model¹⁵ was used.

6 Summary

Generic adversary emulation plans including cloud platform specifics, elastic search data filters, dashboards, configuration, and Terraform scripts were all created through the course of this work. Creating effective analytics for different cloud environments proved challenging for the team, as well as gaps in cloud logging that limited analytic effectiveness. Sharing cloud analytics within the community is encouraged. Such sharing allows defenders to learn from each other and build upon the lessons learned and experience of others. The Analytics identified through the Center's Cloud Analytics project will be proposed as Sigma rules, as appropriate. Future opportunities related to this work include creating a Sigma converter to convert Sigma rules to Google Big Query. For this work, pysigma seems to be the preferred solution versus sigmac. The Center staff have ongoing discussions with the Sigma maintainers and will continue work to coordinate such efforts. Another potential follow-on research opportunity would include enhancements to Elastic Search ECS common data structure to allow for expanded mapping scope.

7 Appendix A – Terminology

Term	Description
Engenuity 16	MITRE Engenuity works in collaboration with the private sector to strengthen our critical infrastructure.

Center for Threat Informed Defense ¹⁷	The Center is a non-profit, privately funded research and development organization operated by MITRE Engenuity. The Center's mission is to advance the state of the art and the state of the practice in threat-informed defense globally.
Cyber Analytics Repository (CAR) 18	The MITRE Cyber Analytics Repository (CAR) is a knowledge base of analytics developed by MITRE based on the MITRE ATT&CK adversary model.
MITRE ATT&CK 19	MITRE ATT&CK is a globally accessible knowledge base of adversary tactics and techniques based on real-world observations.
ATT&CK Cloud ²⁰	ATT&CK Cloud corresponds to tactics and techniques representing the MITRE ATT&CK Matrix for Enterprise covering cloud-based techniques.
Sigma ²¹	Sigma is a generic and open signature format that allows you to describe relevant log events in a straightforward manner. The rule format is very flexible, easy to write and applicable to any type of log file.
ATT&CK Evaluations ²²	The ATT&CK Evaluations program, operated by MITRE Engenuity, provides open and fair evaluations based on ATT&CK.
Infrastructure as a Service (IaaS) ²³	Infrastructure as a Service (IaaS) are online services that provide high-level APIs used dereference various low-level details of underlying network infrastructure like physical computing resources, location, data partitioning, scaling, security, and backup.
Infrastructure as Code (laC) 24	Infrastructure as Code (IaC) is the process of managing and provisioning computer data centers through machine-readable definition files, rather than physical hardware configuration or interactive configuration tools. Software examples include Terraform, Bicep, and CloudFormation.



8 Appendix B - Cloud Offensive Toolkits

8.1 Azure

- MicroBurst <u>GitHub NetSPI/MicroBurst: A collection of scripts for assessing Microsoft</u>
 Azure security
- PowerZure <u>GitHub hausec/PowerZure</u>: <u>PowerShell framework to assess Azure</u> <u>security</u>

8.2 AWS

- weirdAAL GitHub carnalOwnage/weirdAAL: WeirdAAL (AWS Attack Library)
- pacu <u>GitHub RhinoSecurityLabs/pacu: The AWS exploitation framework, designed for testing the security of Amazon Web Services environments.</u>

8.3 GCP

- gcpHound gcpHound : A Swiss Army Knife Offensive Toolkit for Google Cloud Platform (GCP)
- GCPBucketBrute <u>GitHub RhinoSecurityLabs/GCPBucketBrute</u>: A script to enumerate Google Storage buckets, determine what access you have to them, and determine if they can be privilege escalated.

¹⁰ "Google Cloud Platform Acceptable Use Policy," [Online]. Available: https://cloud.google.com/terms/aup/ [Accessed 31 March 2022].

¹¹ "Google Cloud Platform Terms of Service," [Online]. Available: https://cloud.google.com/terms/. [Accessed 31 March 2022].

¹² Microsoft Cloud, "Penetration Testing Rules of Engagement," [Online]. Available: https://www.microsoft.com/en-us/msrc/pentest-rules-of-engagement?rtc=1. [Accessed 31 March 2022].

¹³ Adversary Emulation Library. [Online]. Available: https://github.com/center-for-threat-informed-defense/adversary emulation library/. [Accessed 12 April 2022].

¹⁴ CALDERA Plugin: Emu. [Online]. Available: https://github.com/mitre/emu. [Accessed 12 April 2022].

- 15 Assume breach Cybersecurity Attack and Defense Strategies [Book] (oreilly.com)
- ¹⁶ Engenuity, https://mitre-engenuity.org/
- ¹⁷ Center for Threat Informed Defense, https://ctid.mitre-engenuity.org/
- ¹⁸ Cyber Analytics Repository (CAR), https://car.mitre.org/
- ¹⁹ MITRE ATT&CK, https://attack.mitre.org/
- ²⁰ ATT&CK Cloud, https://attack.mitre.org/matrices/enterprise/cloud/
- ²¹ Sigma, [Online]. Available: https://github.com/SigmaHQ/sigma. [Accessed 31 March 2022].
- ²² ATT&CK Evaluations, https://attackevals.mitre-engenuity.org/index.html
- ²³ Infrastructure as a Service (laaS), https://en.wikipedia.org/wiki/Infrastructure as a service
- ²⁴ Infrastructure as Code (IaC), https://en.wikipedia.org/wiki/Infrastructure as code

¹ Specification: Sigma Correlations · SigmaHQ/sigma Wiki (github.com)

² Compare AWS and Azure services to Google Cloud | Google Cloud Free Program

⁴ Matrix - Enterprise | MITRE ATT&CK®

⁵ Configure anonymous public read access for containers and blobs - Azure Storage | Microsoft Docs

⁶ https://cloud.google.com/storage/docs/access-control/making-data-public

⁷ Palacin and Safari, *Practical Threat Intelligence and Data-Driven Threat Hunting*.

⁸ sigma - Visual Studio Marketplace

⁹ "AWS Customer Support Policy for Penetration Testing," [Online]. Available: https://aws.amazon.com/security/penetration-testing/. [Accessed 31 March 2022].