

Serverless Data Processing (CSCI 5410)

Dr. Saurabh Dey

Outline

1. Cloud Storage
2. Cloud Security



This content is protected and may not be shared, uploaded, or distributed

What is Cloud Storage?

- A cloud storage is used to store data on remote space or servers that can be accessed over internet.
- Users either pay or use free tier service, which is responsible for management, security, maintenance, and backup of the data
- Data centers with massive servers are used to store the data



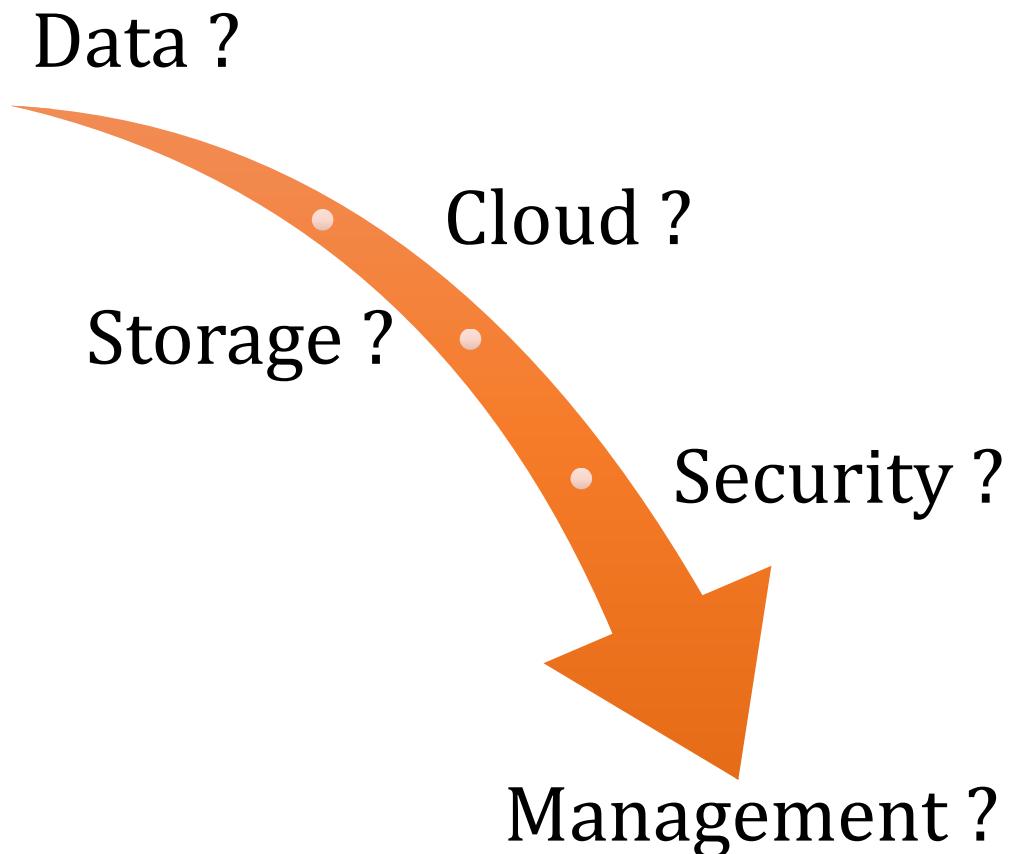
This content is protected and may not be shared, uploaded, or distributed.

Citation: https://www.alibabacloud.com/blog/what-is-cloud-storage-and-how-does-it-work_595461



- Data centers are placed in different continents to ensure efficient, and reliable service.
- Multiple file types, and different file sizes are supported by the storage services
- Uses pay-per-use model.
- Data can be securely accessed with the help of multilayer authentication distributed

Storing Data in the Cloud



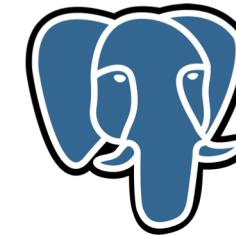
What are the types of Data?

Structured Data

- Social Insurance Number (SIN)
- Date of Birth
- Employment Record
- Address

Unstructured Data

- Images
- Videos
- Email Messages
- Audio
- Reports



Structured Data needs RDBMS

- Storing relational data with transaction management capabilities
- Advanced Querying facilities

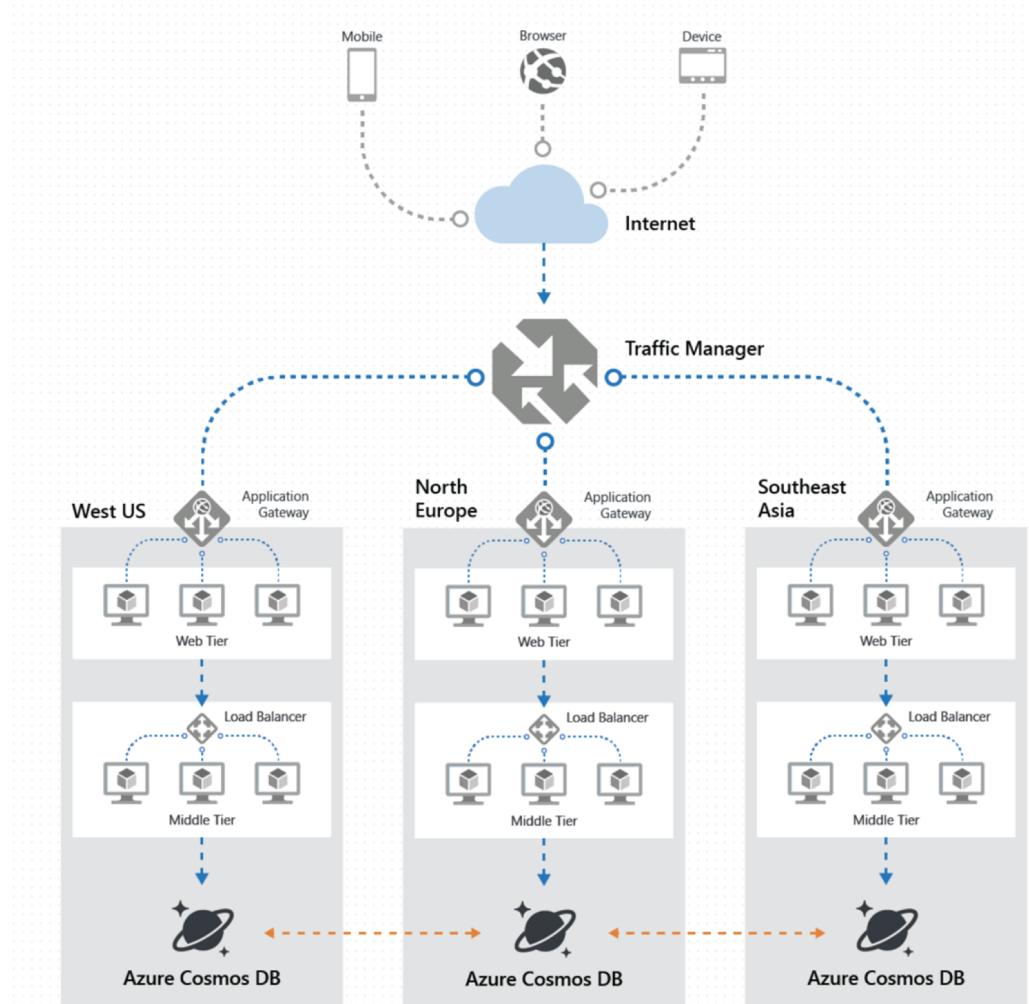
Azure SQL Database, Azure Database for PostgreSQL, and MySQL

AWS RDS manages relational database service

GCP – Cloud SQL, and Cloud Spanner

How to store Unstructured Data?

- Storing and managing unstructured data in cloud requires a different strategy.
- NoSQL database allows flexible schema, and easy development.



This content is protected and may not be shared, uploaded, or
distributed

Azure Cosmos DB – Low latency, and high availability

AWS NoSQL Database

- AWS has Dynamo DB, which is very popular NoSQL database.
- With DynamoDB scale up and scale down is possible of tables' throughput capacity without performance degradation.
- Working with DynamoDB requires simple API operations.
 1. Control Plane
 2. Data Plane
 3. DynamoDB Streams
 4. Transactions

The screenshot shows the AWS DynamoDB console interface. At the top, there's a navigation bar with links like 'Services', 'Resource Groups', and 'Tutorial'. Below that is a 'Create DynamoDB table' form where 'Table name' is set to 'Test' and 'Primary key' is set to 'testClass' (String type). Under 'Table settings', the 'Use default settings' checkbox is checked, and a list of default configurations is shown. The main view then transitions to the 'Overview' tab of the 'Test' table's details page. It displays basic information such as 'Stream enabled: No', 'View type: -', and 'Latest stream ARN: -'. A 'Manage Stream' button is available. The 'Table details' section provides a comprehensive list of table properties, including:

- Table name: Test
- Primary partition key: testClass (String)
- Primary sort key: -
- Point-in-time recovery: DISABLED [Enable](#)
- Encryption Type: DEFAULT [Manage Encryption](#)
- KMS Master Key ARN: Not Applicable
- Encryption Status: Enabled
- CloudWatch Contributor Insights: DISABLED [Manage Contributor Insights](#) NEW
- Time to live attribute: -
- Table status: Active
- Creation date: May 10, 2020 at 10:32:53 PM UTC-3
- Provisioned read capacity mode: Last change to on-demand mode
- Provisioned read capacity units: 5 (Auto Scaling Error)
- Provisioned write capacity mode: Last decrease time
- Provisioned write capacity units: 5 (Auto Scaling Error)

This content is protected and may not be shared, uploaded or distributed

GCP NoSQL/ non-relational database

 Bigtable [← Create an instance](#)

A Cloud Bigtable instance is a container for your clusters. [Learn more](#)

Instance name
For display purposes only

Instance ID
ID is permanent
 Lowercase letters, numbers and hyphens allowed

Storage type ⓘ
Choice is permanent. Applies to all clusters. Affects cost.
 SSD
Lower latency and more rows read per second. Typically used for real-time serving use cases, such as ad serving and mobile app recommendations.
 HDD
Higher latency for random reads. Good performance on scans and typically used for batch analytics, such as machine learning or data mining.

Clusters

Cluster
<input type="text"/> Cluster ID ID is permanent. <input type="text"/> Lowercase letters, numbers and hyphens allowed

Cost estimate

Monthly resource costs
Monthly costs reflect Bigtable resources only. Network traffic (replication and internet egress) costs are dependent on the location of your clusters and application request behaviour. [Learn more](#)

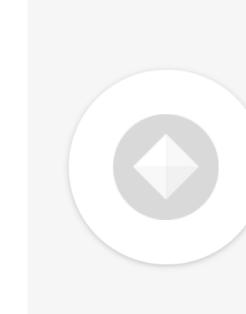
Try another storage size (per cluster)
 GB

Item	Estimated cost
1 cluster	\$474.50/month
1 GB HDD	\$0.03/month
Total	\$474.53/month

Node charges are for provisioned resources, regardless of node usage. The same node charges apply even if your instance is inactive. [Learn more](#)

Summary
Monthly charge: \$474.53 per month (1 GB data, 1 nodes)
Effective hourly rate: \$0.65

Cloud Bigtable



Google Cloud Memorystore for Redis API

Google

Creates and manages Redis instances on the Google Cloud Platform.

[ENABLE](#)

[TRY THIS API](#)

Type
[APIs & services](#)

Last updated
30/11/2018, 23:44

Service name
redis.googleapis.com

Overview

Creates and manages Redis instances on the Google Cloud Platform.

About Google

Google's mission is to organize the world's information and make it universally accessible and useful. Through products and platforms like Search, Maps, Gmail, Android, Google Play, Chrome and YouTube, Google plays a meaningful role in the daily lives of billions of people.

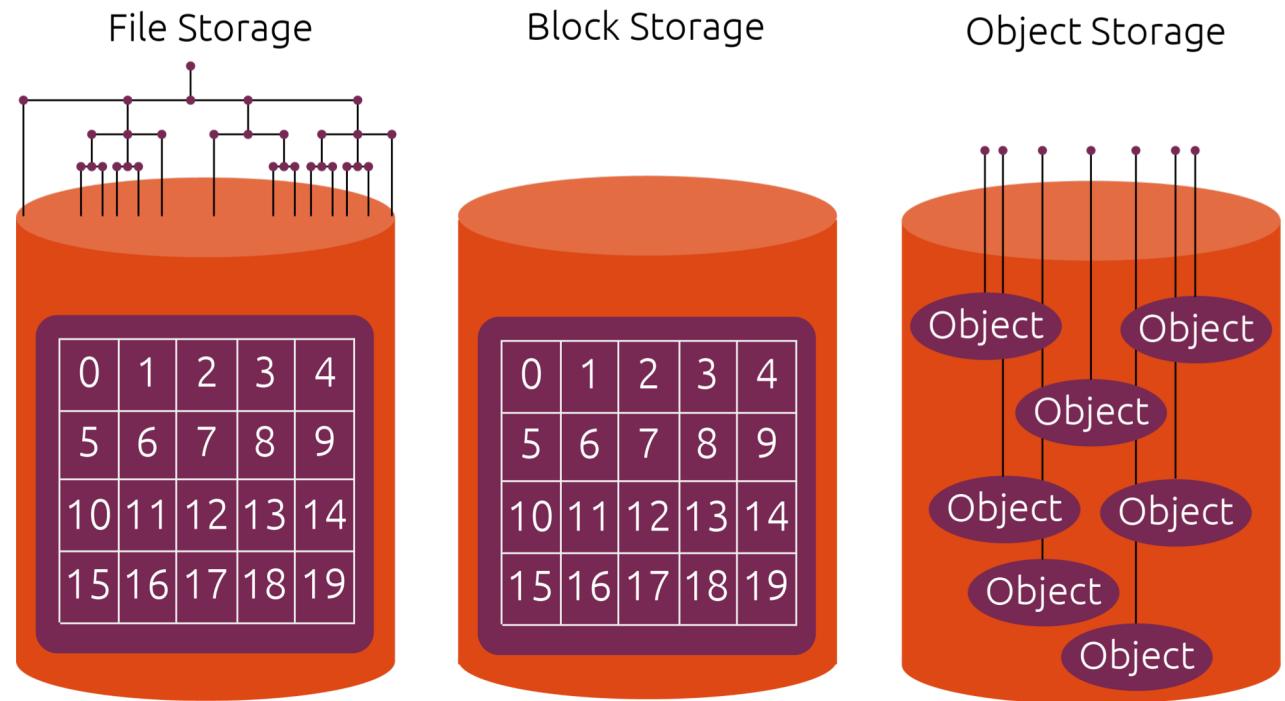
This content is protected and may not be shared, uploaded, or distributed

<https://cloud.google.com/products/databases>

Memorystore

Is it important to study Cloud Storage?

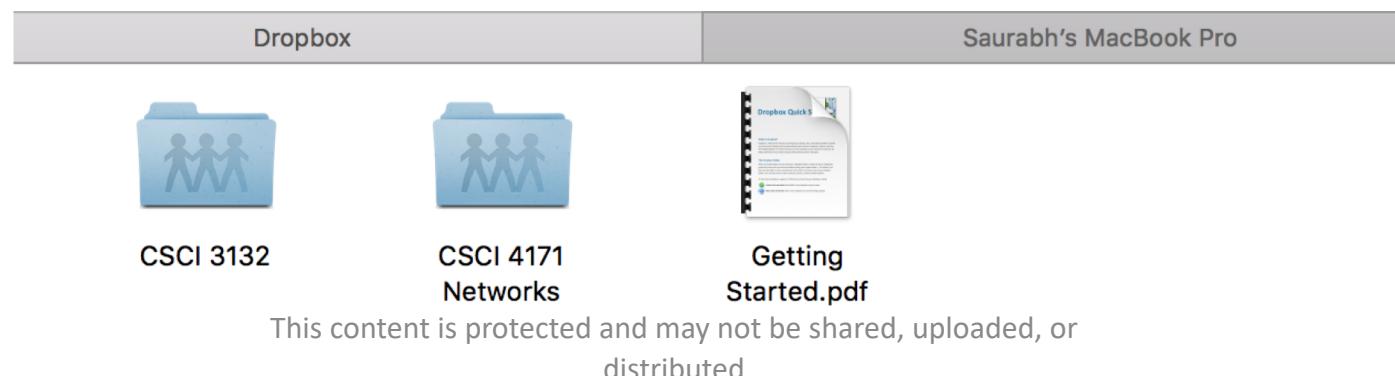
- Block, File, Object storage are broad categories of data storage
- Storage is accessed by Web service API
- Block storage is raw storage and it is usually faster
- For unstructured data and searching capability object storage is better



<https://ubuntu.com/blog/what-are-the-different-types-of-storage-block-object-and-file>

Provisioning of Cloud Storage

- There are two major classes of storage: unmanaged and managed storage.
- In unmanaged storage, the storage service provider makes storage capacity available to users
- The provider defines the nature of the storage, how it may be used, and by what applications.
- In an unmanaged cloud storage service, disk space is made available to users as a sized partition.



Provisioning of Cloud Storage

- In a managed cloud storage system, the user provisions storage on demand and pays for the storage using a pay-as-you-go model.
- In this kind of storage, the user partition the raw disk, and format it according to the requirement.



S3
Blob
Cloud Storage

This content is protected and may not be shared, uploaded, or

<https://aws.amazon.com/s3/>

Cloud Computing Security

This content is protected and may not be shared, uploaded, or
distributed

Security

- Malware can travel across the globe
- Compromised critical information infrastructure can cause damages at large scale.
- Cyberwarfare is a new term in this game.
- Data is more vulnerable in storage, as it is kept in storage for extended periods of time.
- We need to assess the security requirements.

Citation:

[1] Dan C. Marinescu. (2018) *Cloud Computing Theory and Practice*, Second Ed

[2] Holger Schulze, "Cloud Security Report 2019", Cyber Security Insiders



42%

Unauthorized
access



42%

Insecure interfaces
/APIs

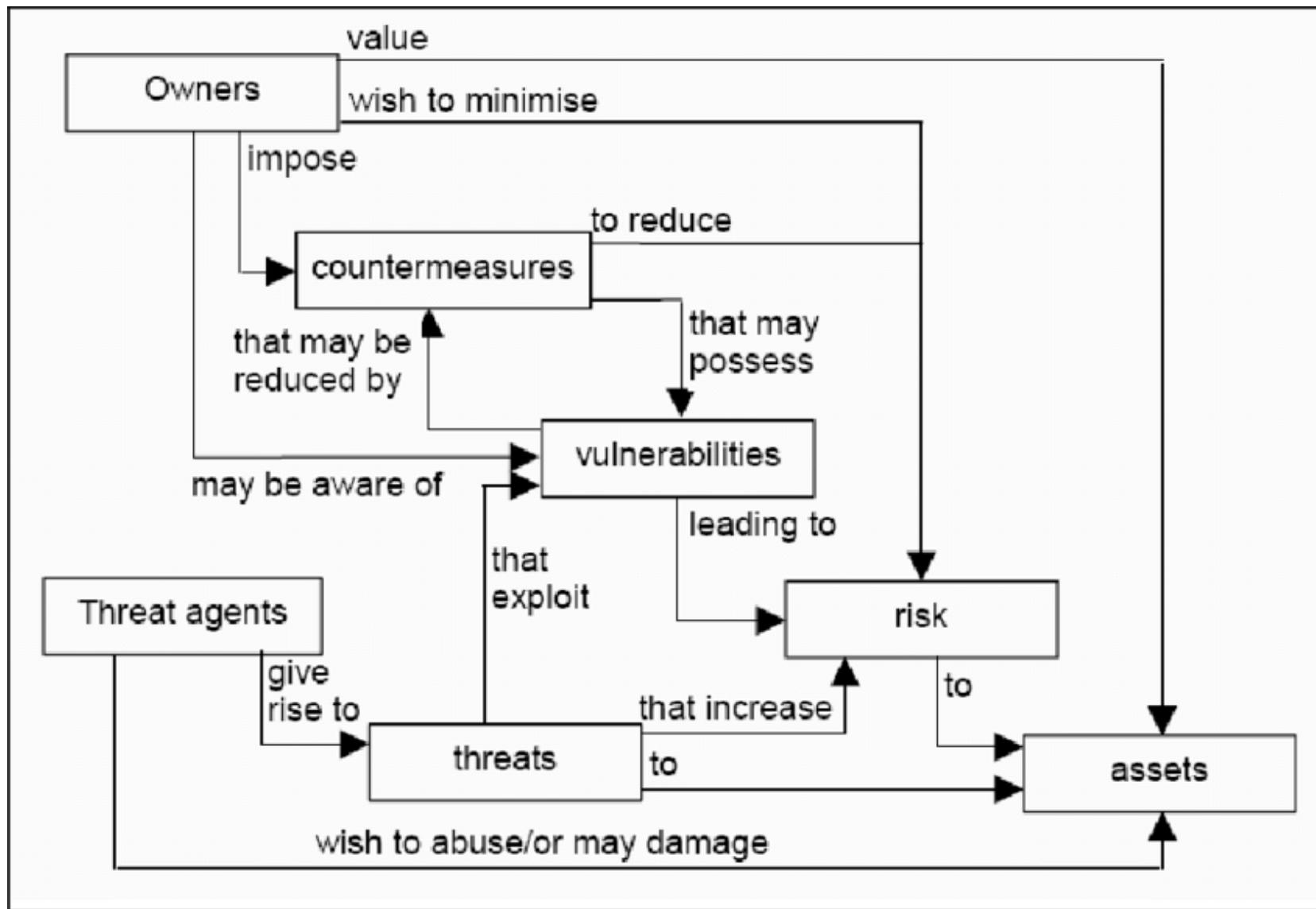


40%

Misconfiguration of
the cloud platform
/wrong setup

This content is protected and may not be shared, uploaded, or distributed

Security Model



- 1. Owners:** Cloud Provider, Service Owner, cloud consumer
- 2. Threat agents:** An entity that poses a threat because it is capable of carrying out an attack
- 3. Threats:** These make the vulnerability happen
- 4. Vulnerabilities:** possibility of being attacked
- 5. Risk:** possible threat to the security of assets
- 6. Assets:** IT resources, Databases, data
- 7. Countermeasures:** Security measures

ISO 15408 Risk Context

This content is protected and may not be shared, uploaded, or

Source: Floarea, Nastase & Pavel, NASTASE. (2007). Risk Management for e-Business. *Informatica Economica Journal*.

Types of Threats

In addition to Denial of Service (DoS), or Man-in-the-middle or malicious insiders

Virtualization Attack

- Authorized users with administrative access level may abuse the access to launch attack on the physical hardware

Insufficient Authorization

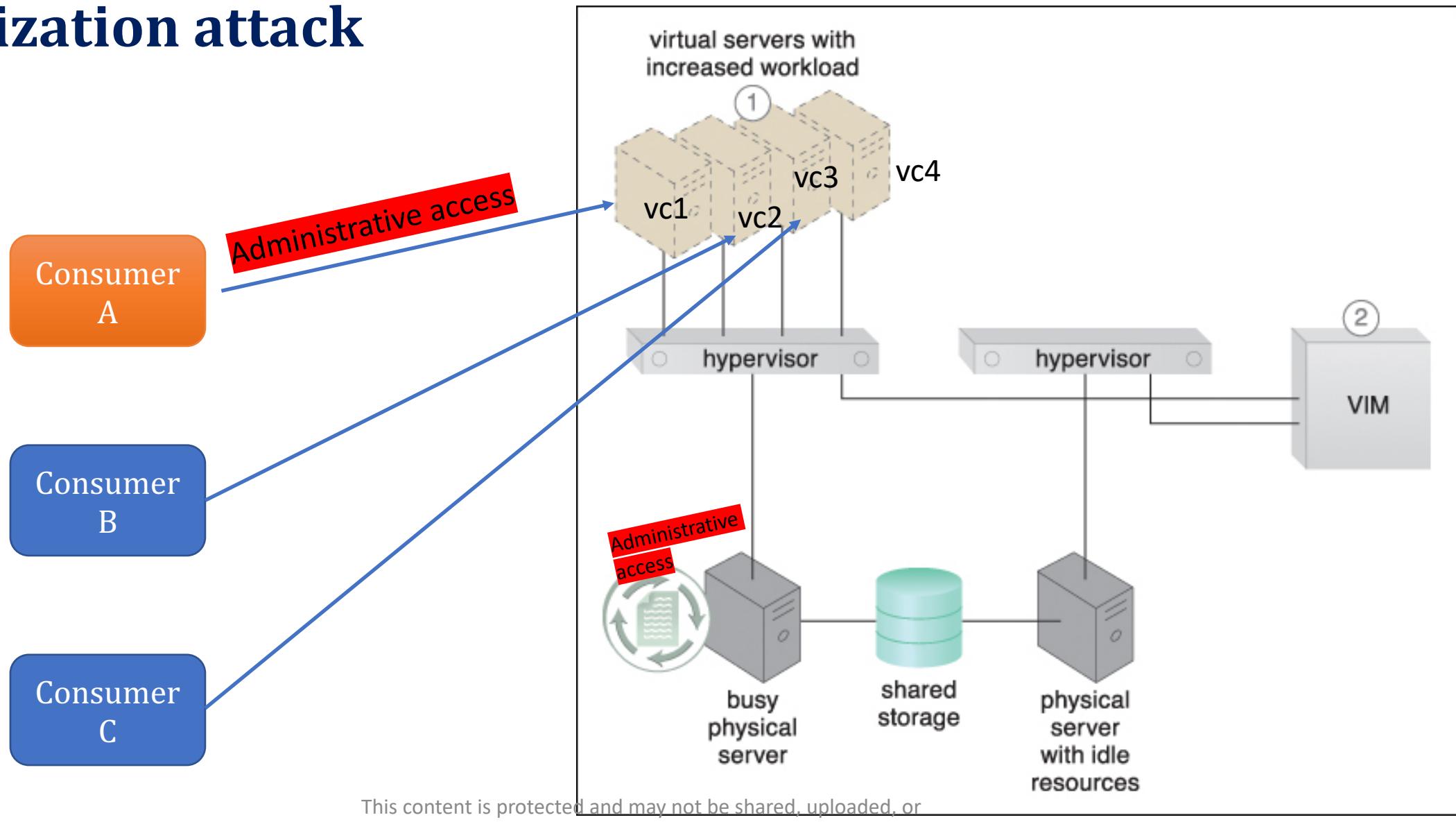
- The insufficient authorization occurs when access is granted to an attacker too broadly

Overlapping Trust Boundaries

- Malicious cloud service consumers can target shared IT resources with the intention of compromising IT resources that share the same trust boundary

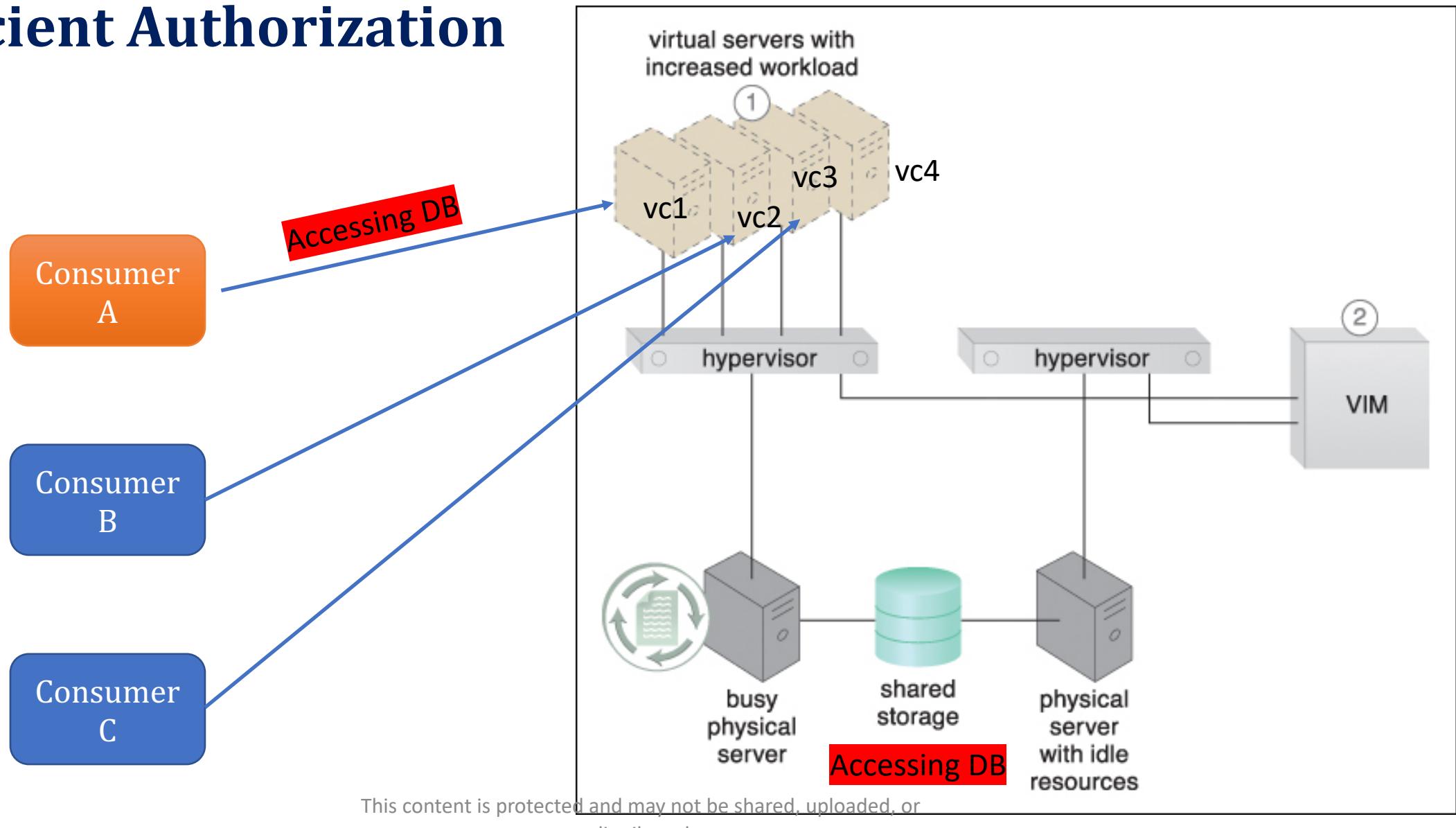
Understand the Threats

Virtualization attack

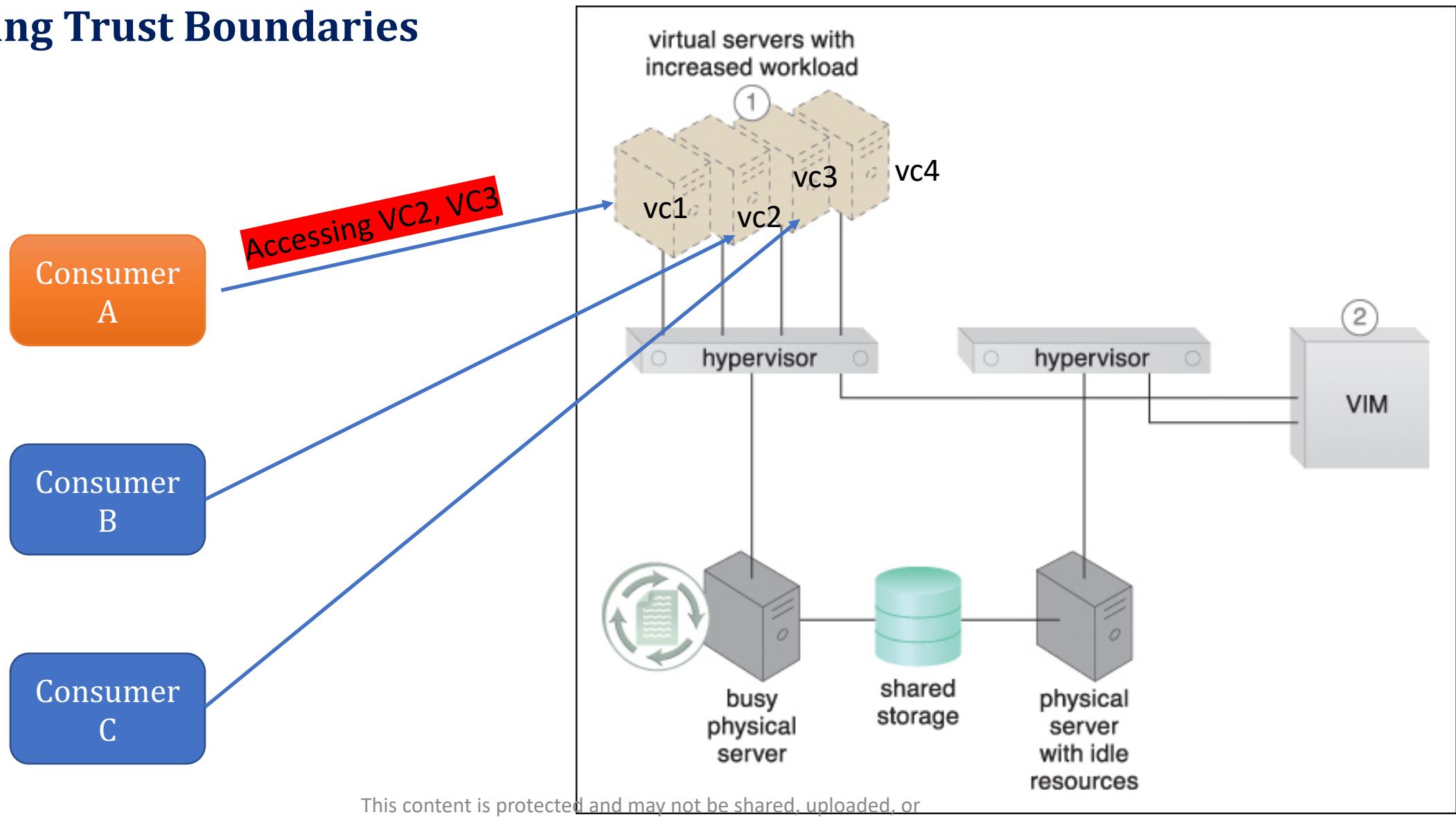


This content is protected and may not be shared, uploaded, or distributed

Understand the Threats Insufficient Authorization



Understand the Threats Overlapping Trust Boundaries

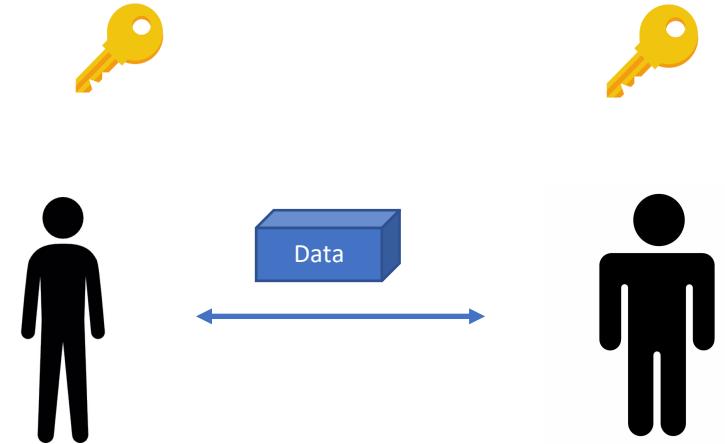


CSA (Cloud Security Alliance)

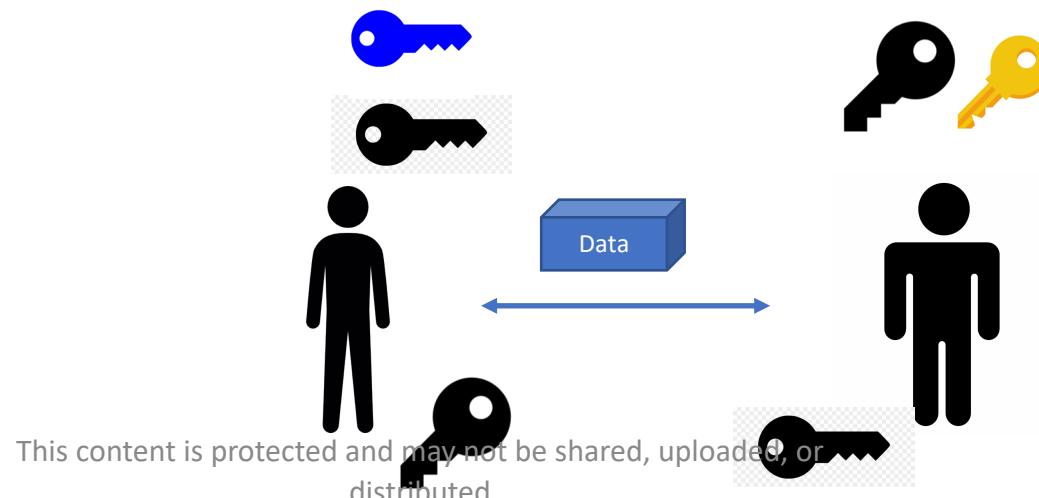
- Member driven organization formed in 2008
- Membership is comprised of many of the industry's large scale vendors and suppliers
- CSA provides guidelines and best practices organized into 14 domains
 - Cloud Architecture, Compliance and Audit, Data Security, Portability, interoperability etc.
- Cloud Control Matrix (CCM) – provides a security control list and framework that enables detailed understanding of security concepts and principles.

Encryption

- Symmetric encryption uses single key
Blowfish, AES, DES etc.



- Asymmetric encryption uses two keys {private, public}
ECC, RSA



Questions to Consider

- Between GCP Bigtable and Memorystore, what will you select for a car manufacturing plant?
- Is Object storage better than Block Storage?
- When do you prefer symmetric encryption over asymmetric in your cloud application?



Thank You



This content is protected and may not be shared, uploaded, or
distributed