Serverless Data Processing (CSCI 5410)

Dr. Saurabh Dey

Outline

1. Cloud Storage
2. Cloud Security

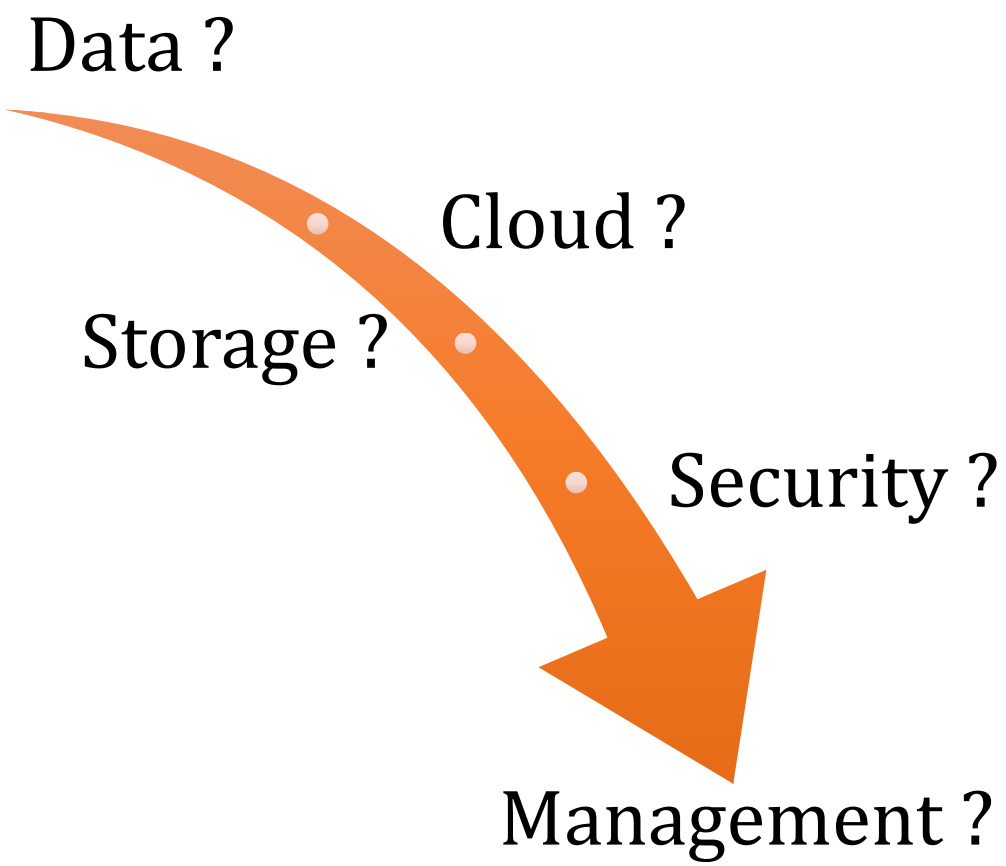
What is Cloud Storage?

- 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

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

Storing Data in the Cloud



What are the types of Data?

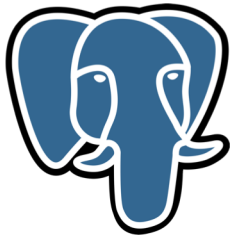
Structured Data

- 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

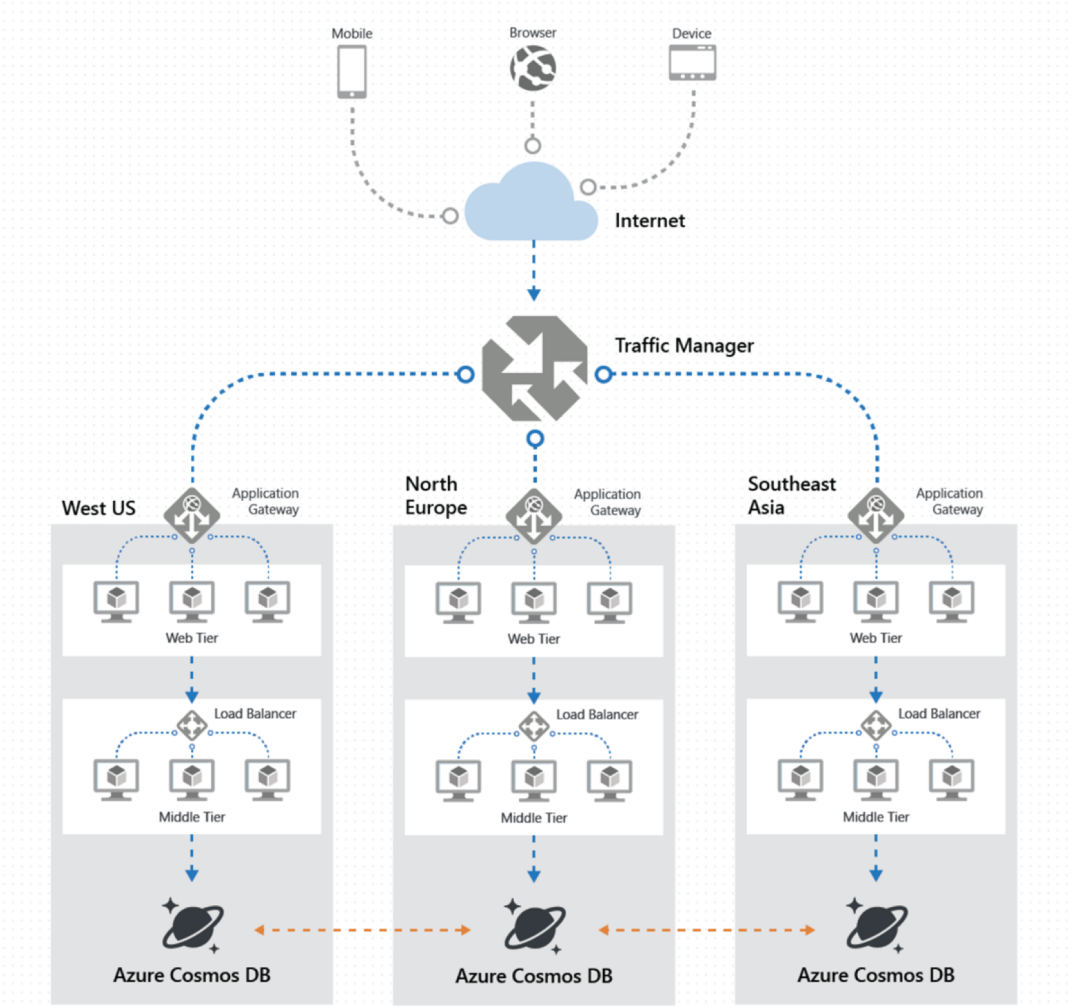
AWS RDS manages relational database service

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.

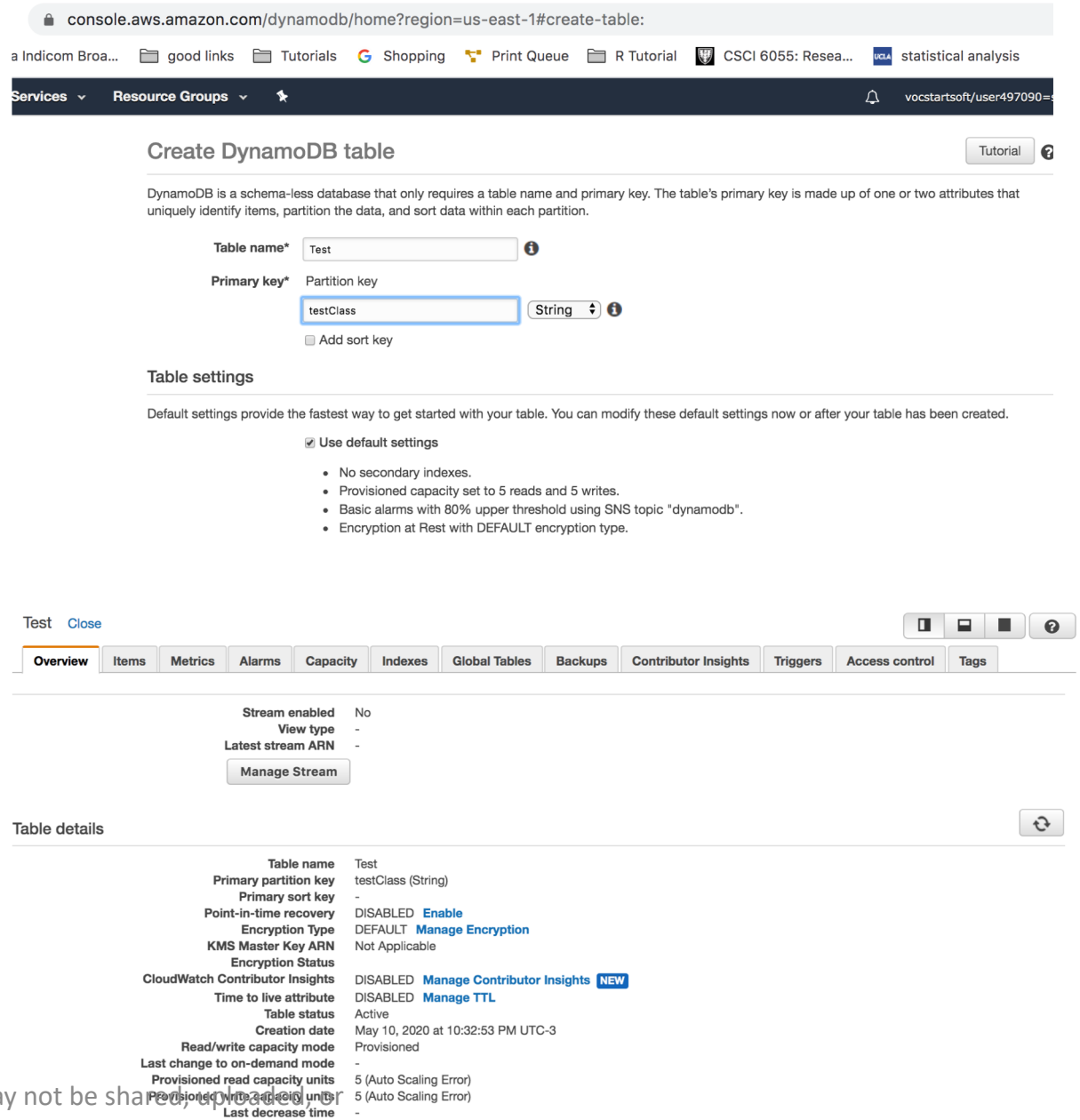


Azure Cosmos DB – Low latency, and high availability

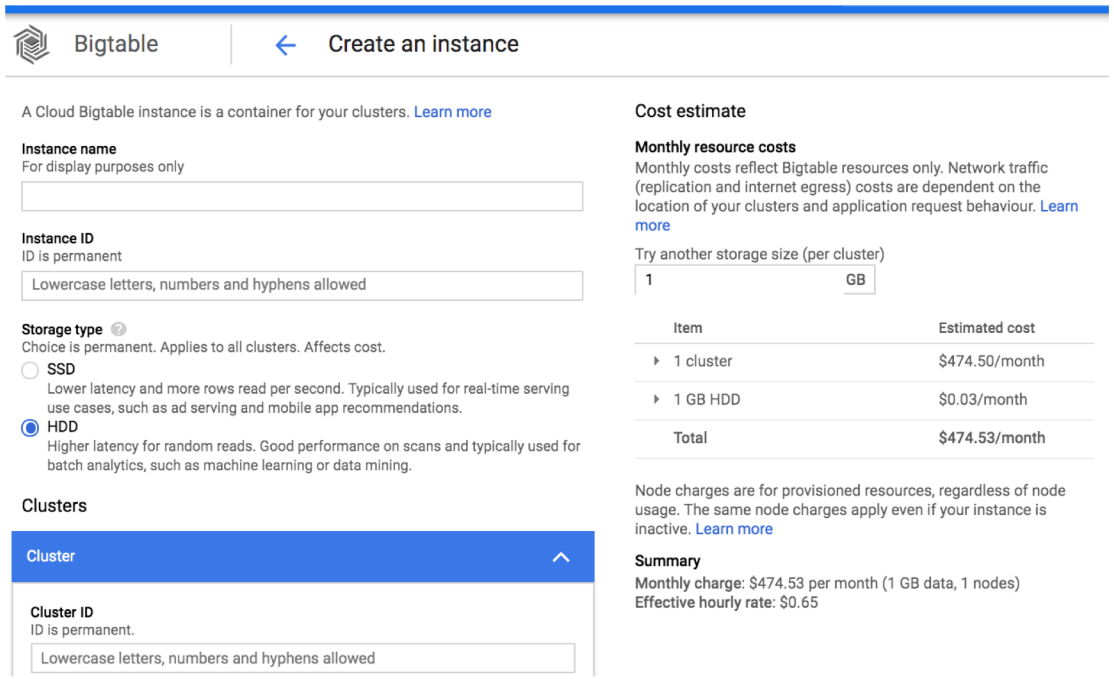
https://docs.microsoft.com/en-ca/azure/cosmos-db/distribute-data-globally

AWS NoSQL Database

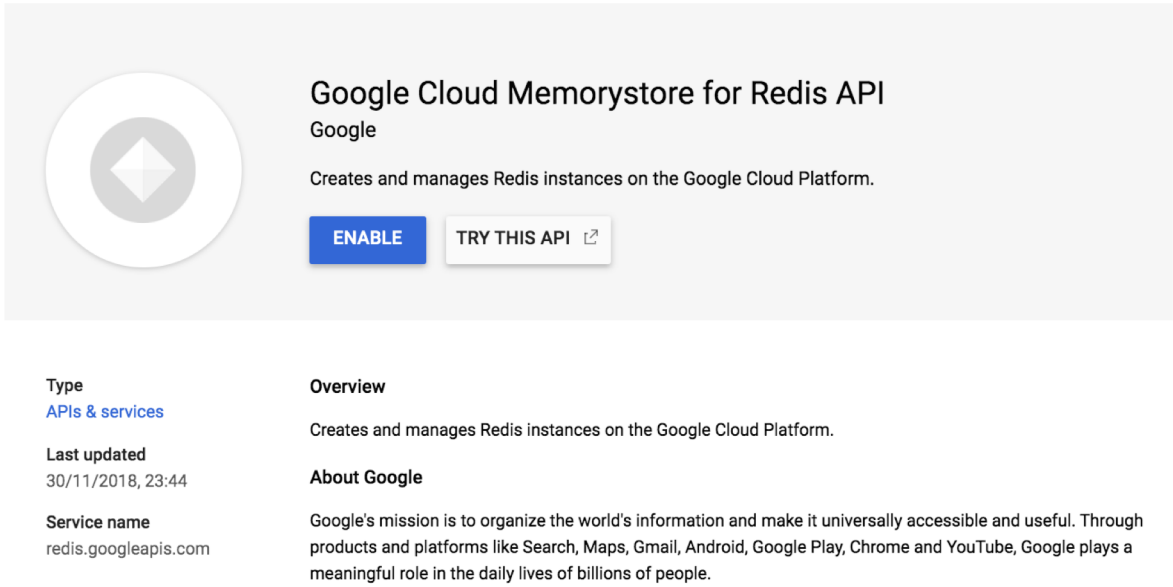
- AWS has Dynamo DB, which is very popular NoSQL database.
- With DynamoDB scale up and scale down is possible of tables' throughput capacity without performance degradation.
- Working with DynamoDB requires simple API operations.
    1. Control Plane
    2. Data Plane
    3. DynamoDB Streams
    4.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



GCP NoSQL/ non-relational database



Cloud Bigtable

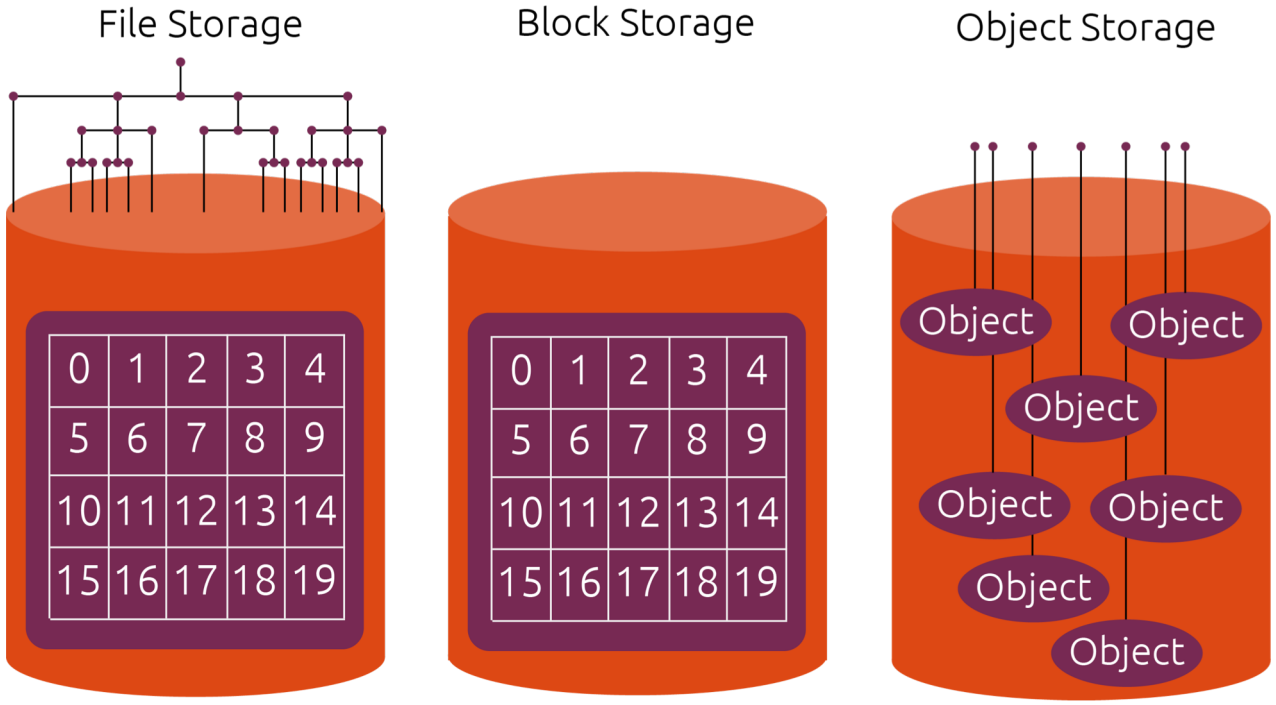


Memorystore

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

Is it important to study Cloud Storage?

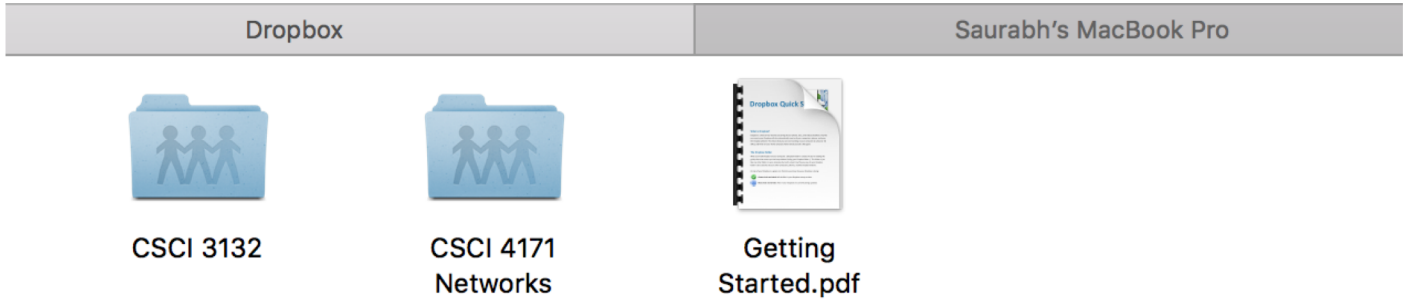
- Block, File, Object storage are broad categories of data storage
- Storage is accessed by Web service API
- Block storage is raw storage and it is usually faster
- For unstructured data and searching capability object storage is better



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

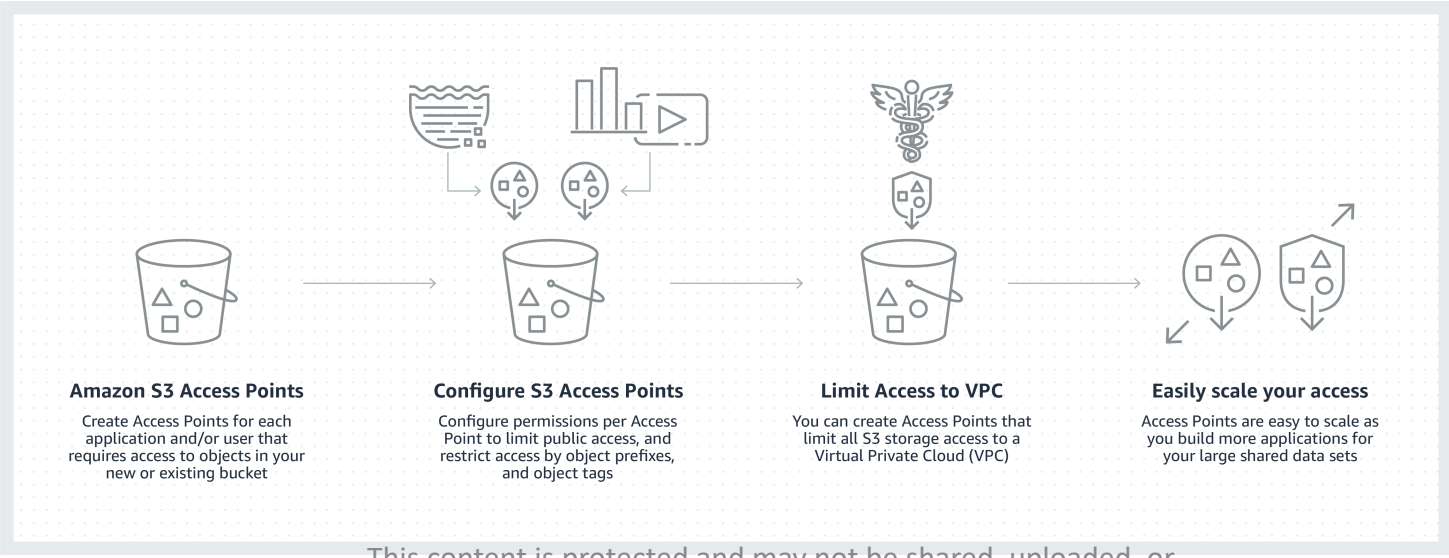
Provisioning of Cloud Storage

- 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.



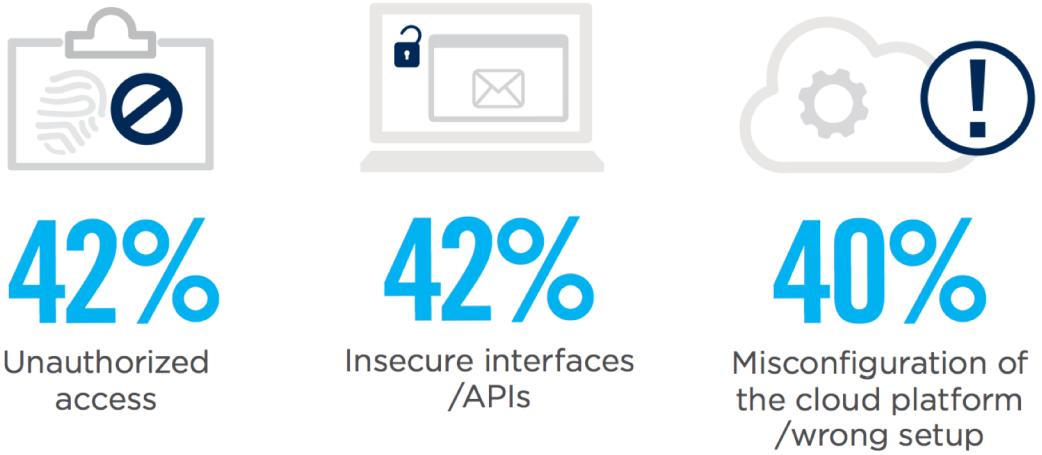
S3
Blob
Cloud Storage

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

Cloud Computing Security

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.

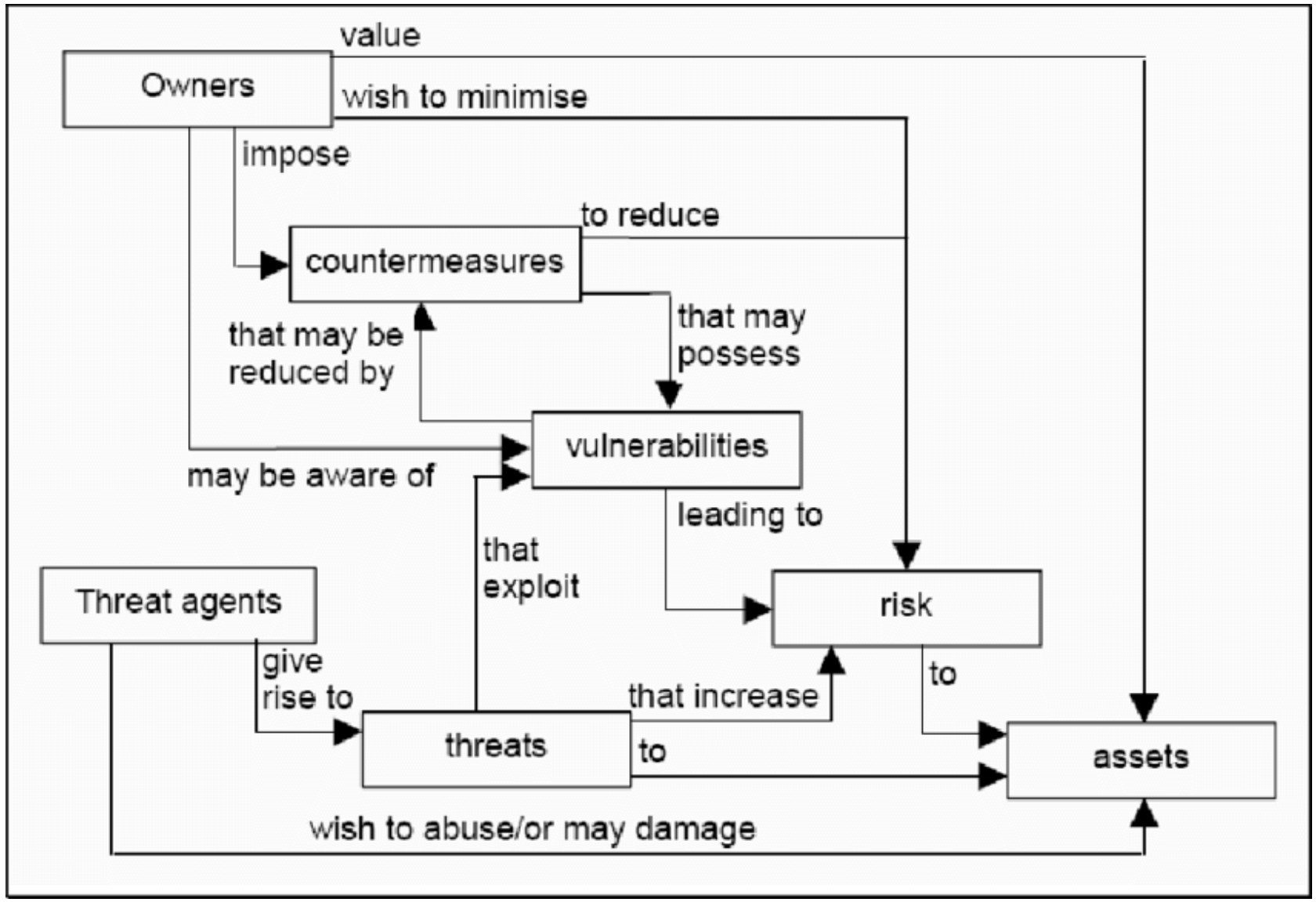


Citation:

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

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

Security Model



ISO 15408 Risk Context

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

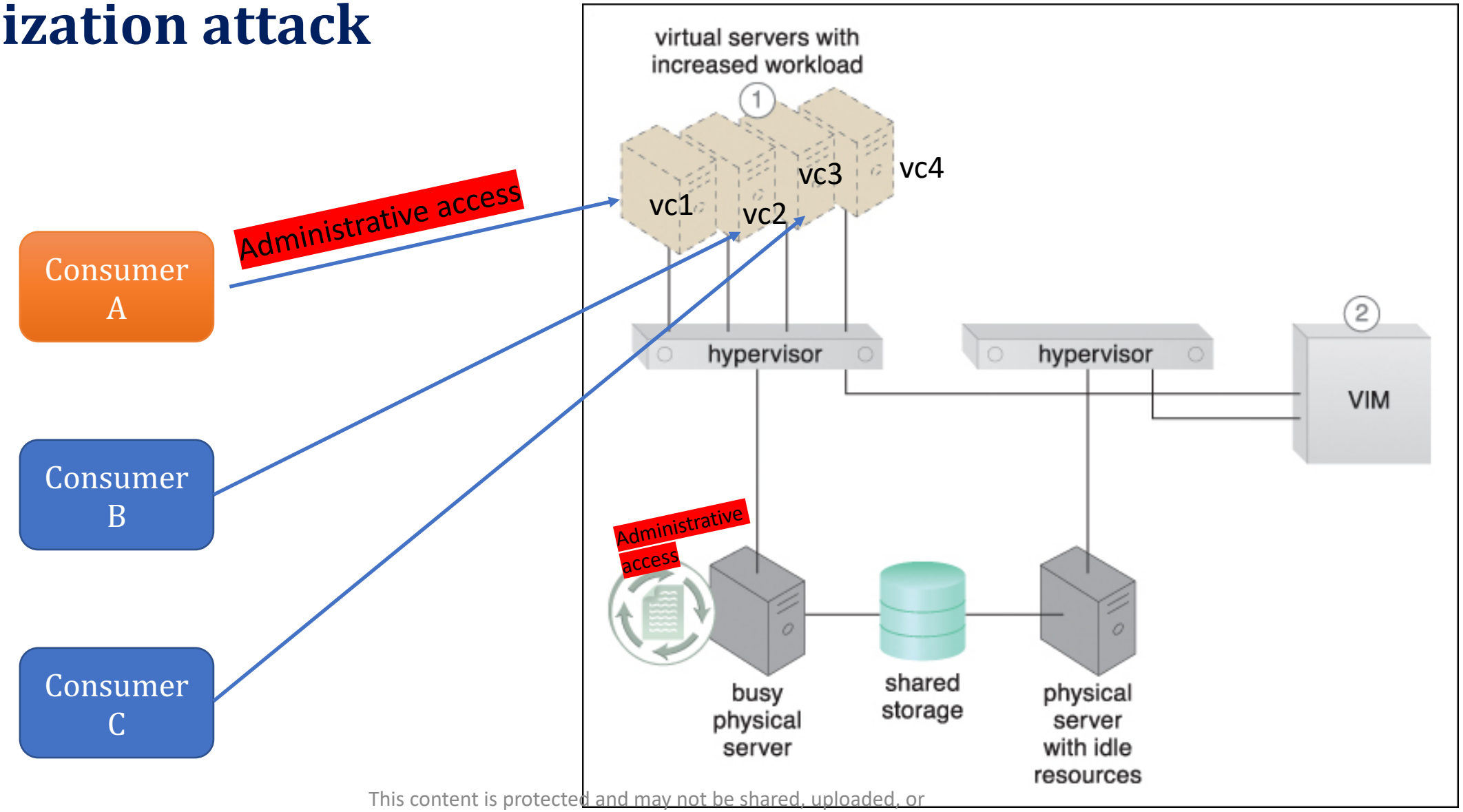
Source: Floarea, Nastase & Pavel, NASTASE. (2007). Risk Management for e-Business. Informatica Economica Journal.

Types of Threats

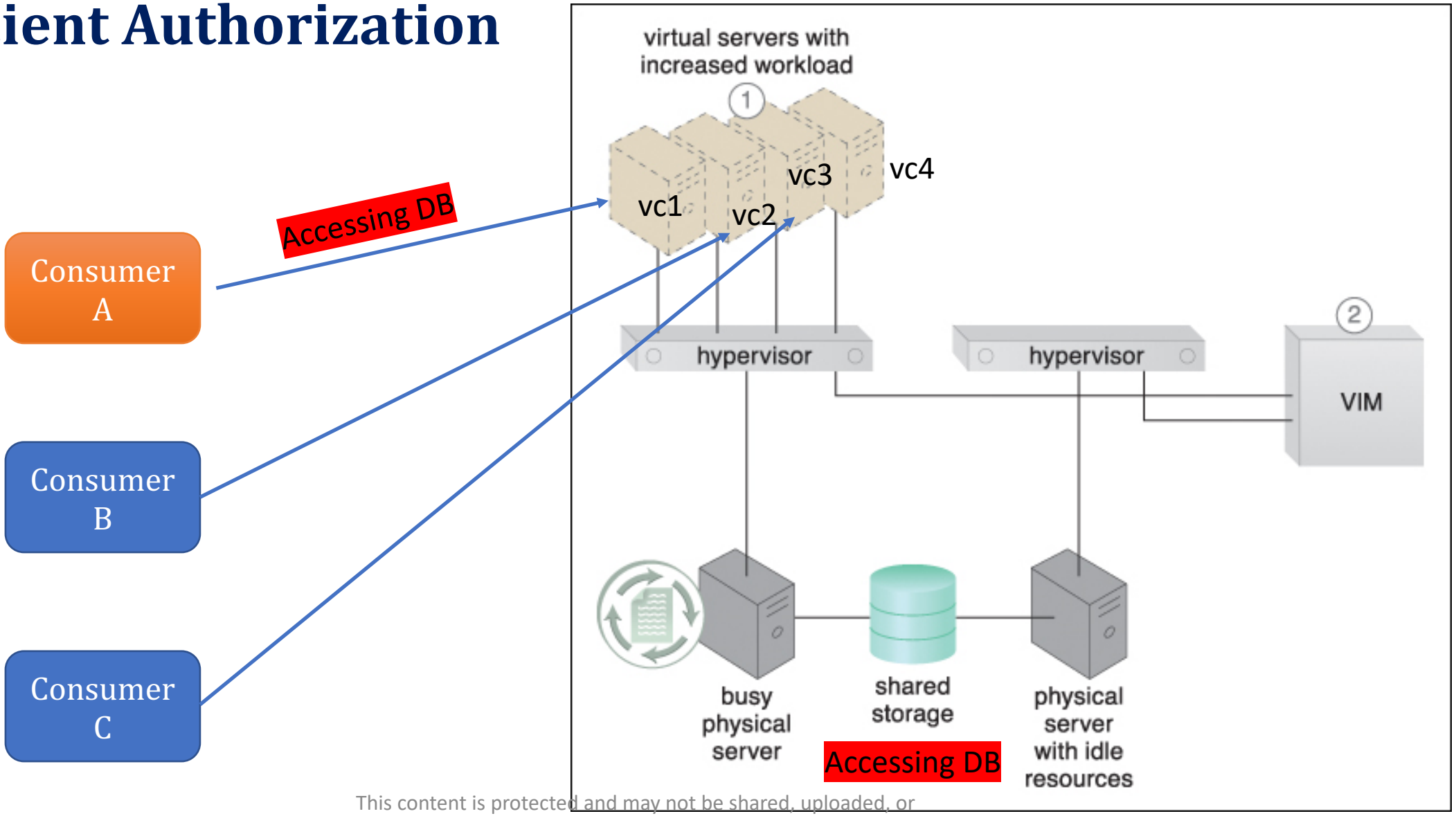
In addition to Denial of Service (DoS), or Man-in-the-middle or malicious insiders

Virtualization Attack	Insufficient Authorization	Overlapping Trust Boundaries
• Authorized users with administrative access level may abuse the access to launch attack on the physical hardware	• The insufficient authorization occurs when access is granted to an attacker too broadly	• Malicious cloud service consumers can target shared IT resources with the intention of compromising IT resources that share the same trust boundary

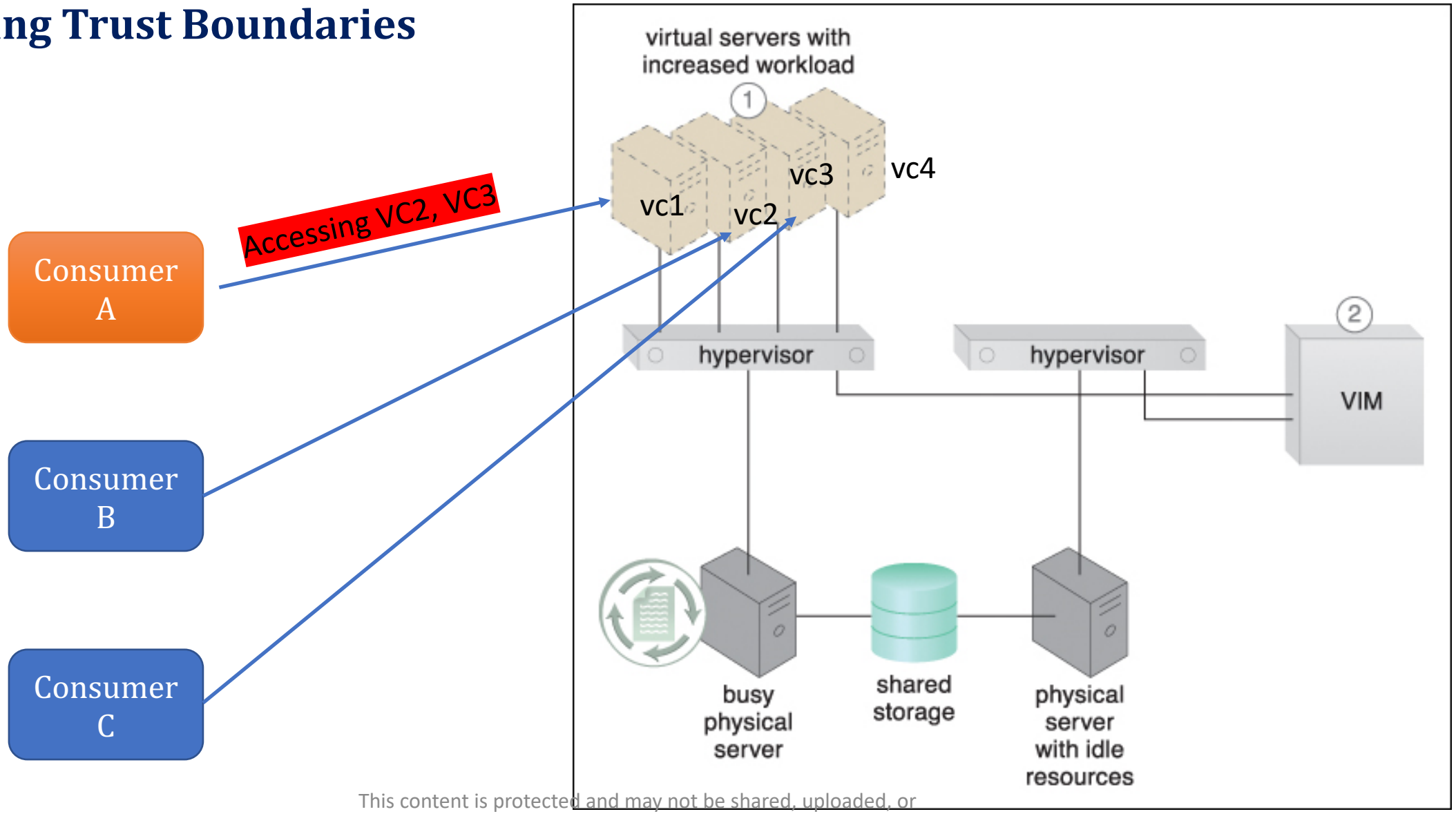
Understand the Threats
Virtualization attack



Understand the Threats
Insufficient Authorization



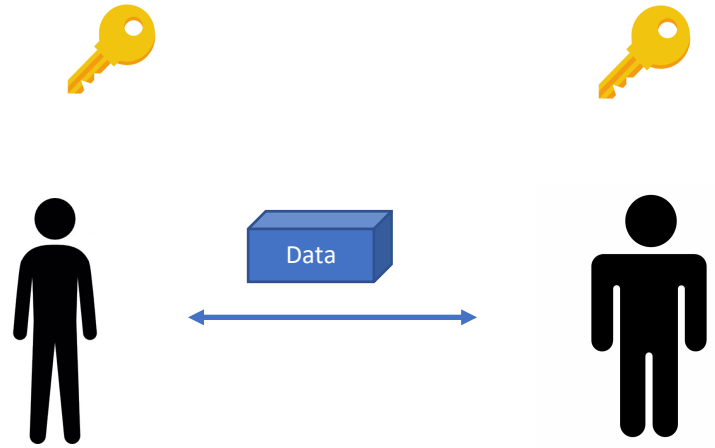
Understand the Threats
Overlapping Trust Boundaries



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

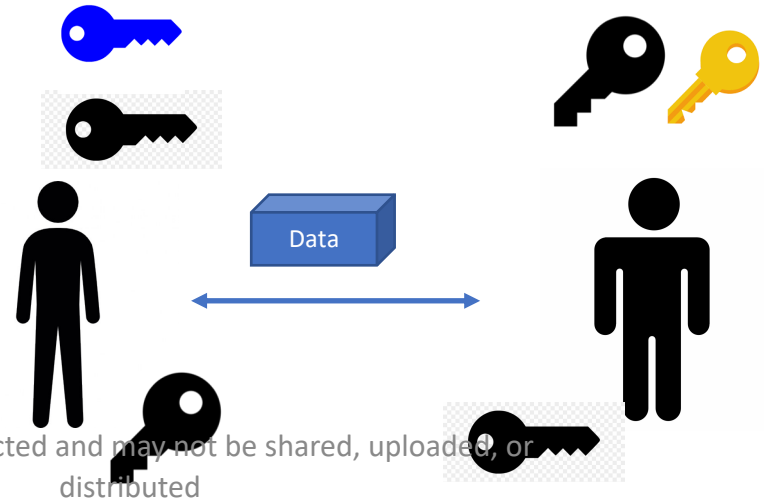


- Symmetric encryption uses single key

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