HPS
https://hps.vi4io.org/

Sonal Lakhotia

Seminar: Newest Trends in High-Performance Data Analytics

Emerging Trends in Cloud Storage

March 2, 2023

Table of contents

1 Introduction

2 Overview of Cloud Data Storage

3 Decentralised Cloud Storage with IPFS

4 Privacy and Security for Cloud Storage

5 Conclusion

What is Cloud Storage?

- Cloud storage is a model of computer data storage in which the digital data is stored in logical pools, said to be on "the cloud".
- Highly fault tolerant through redundancy and distribution of data
- Highly durable through the creation of versioned copies
- Typically eventually consistent with regard to data replicas

Wikipedia, *Cloud Storage*

Importance of Cloud Storage

- Cost effectiveness
- Faster deployment
- Virtually unlimited scalability
- Efficient data management
- Increased agility
- Business continuity

How does Cloud Storage Work?

- Cloud services providers own and operate data storage capacity.
- Cloud storage providers manage data storage aspects such as:
  - Capacity
  - Security
  - Data availability
  - Storage Servers
  - Computing resources
  - Network Data Delivery

1 Introduction

2 Overview of Cloud Data Storage

3 Decentralised Cloud Storage with IPFS

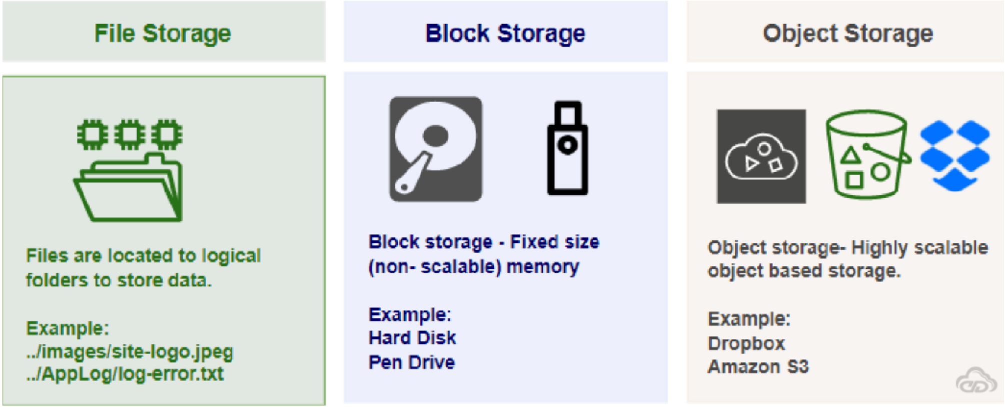
4 Privacy and Security for Cloud Storage

5 Conclusion

Cloud Storage Models

- Public - Data stored in a service provider’s data centers, data can be scaled up or down.
- Private - used by organizations with strict security that can control their data
- Hybrid - A hybrid cloud model comprises private and public cloud storage models.
- Multicloud - More than one cloud model from public or private service providers

Types of Cloud Storage



File Storage

Files are located to logical folders to store data.

Example:
../images/site-logo.jpeg
../AppLog/log-error.txt

Block Storage

Block storage - Fixed size (non- scalable) memory

Example:
Hard Disk
Pen Drive

Object Storage

Object storage- Highly scalable object based storage.

Example:
Dropbox
Amazon S3

AWS, *Types of Cloud Storage*

File Storage

- Stores data in a hierarchical folder and file format.
- File storage is common in personal computing.
- Easy to locate and retrieve individual data items.
- Supported with a Network Attached Storage (NAS) server.
- Examples: Amazon Elastic File System,Qumulo Core.

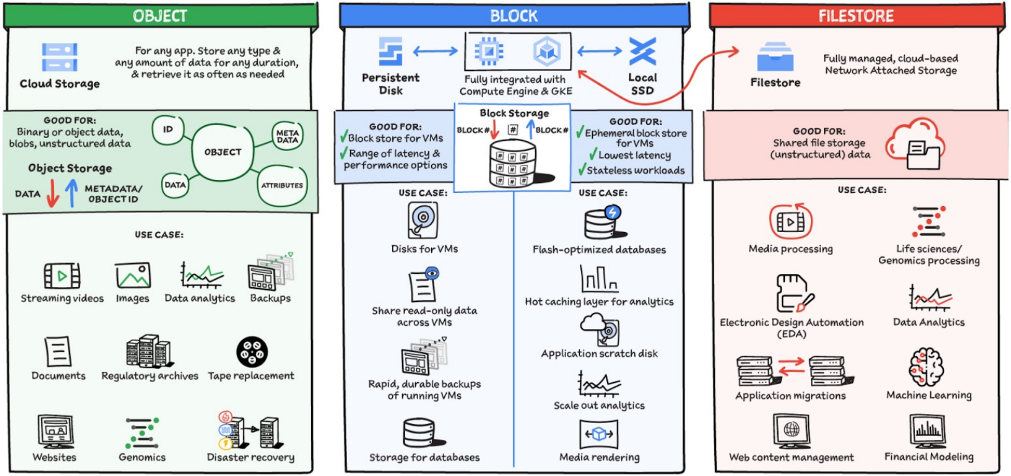
Block Storage

- It offers dedicated, low-latency storage for each host.
- Analogous to direct-attached storage (DAS) or a storage area network (SAN).
- Data is stored as blocks and has its own unique identifier.
- Example: Amazon Elastic Block Store (EBS).

Object Storage

- Data storage architecture for large volumes unstructured data.
- Stores data in the format it arrives in.
- It allows metadata customization for easier data access and analyze.
- Objects are kept in secure buckets that deliver virtually unlimited scalability.
- Amazon S3, Oracle Cloud Storage, Microsoft Azure Storage.

Which Storage to use?



OBJECT

Cloud Storage

For any app. Store any type & any amount of data for any duration, & retrieve it as often as needed

GOOD FOR:
Binary or object data, blobs, unstructured data

Object Storage

DATA ↓ ↑ METADATA/ OBJECT ID

USE CASE:

Streaming videos
Images
Data analytics
Backups
Documents
Regulatory archives
Tape replacement
Websites
Genomics
Disaster recovery

BLOCK

Persistent Disk ⟷ Fully integrated with Compute Engine & GKE ⟷ Local SSD

GOOD FOR:
✓Block store for VMs
✓Range of latency & performance options

Block Storage

GOOD FOR:
✓Ephemeral block store for VMs
✓Lowest latency
✓Stateless workloads

USE CASE:

Disks for VMs
Share read-only data across VMs
Rapid, durable backups of running VMs
Storage for databases

USE CASE:

Flash-optimized databases
Hot caching layer for analytics
Application scratch disk
Scale out analytics
Media rendering

FILESTORE

Filestore

Fully managed, cloud-based Network Attached Storage

GOOD FOR:
Shared file storage (unstructured) data

USE CASE:

Media processing
Life sciences/ Genomics processing
Electronic Design Automation (EDA)
Data Analytics
Application migrations
Machine Learning
Web content management
Financial Modeling

GoogleCloud, *A map of storage options in Google Cloud*

1 Introduction

2 Overview of Cloud Data Storage

3 **Decentralised Cloud Storage with IPFS**

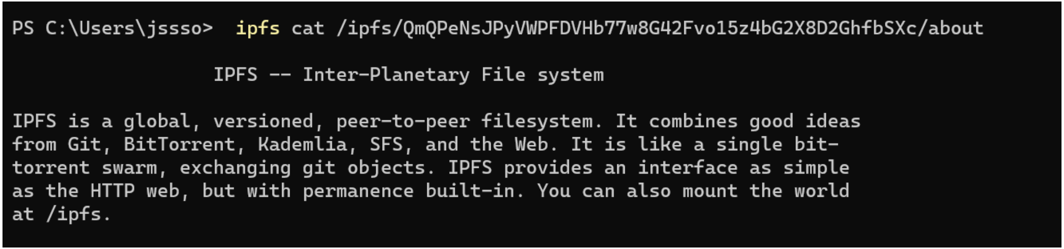
4 Privacy and Security for Cloud Storage

5 Conclusion

Motivation for decentralized data storage

- Centralized cloud storage controls internet traffic.
- Users have less control over data.
- Security and Privacy risks
- Mismanagement of Data
- Monopolisation of costs

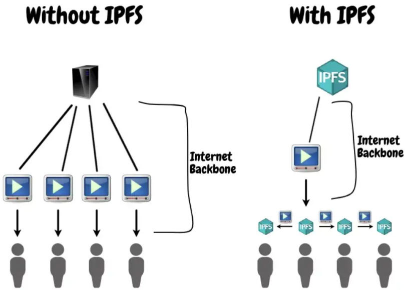
What is IPFS?



```
PS C:\Users\jssso>  ipfs cat /ipfs/QmQPeNsJPyVWPFDVHb77w8G42Fvo15z4bG2X8D2GhfbSXc/about

                  IPFS -- Inter-Planetary File system

IPFS is a global, versioned, peer-to-peer filesystem. It combines good ideas
from Git, BitTorrent, Kademlia, SFS, and the Web. It is like a single bit-
torrent swarm, exchanging git objects. IPFS provides an interface as simple
as the HTTP web, but with permanence built-in. You can also mount the world
at /ipfs.
```

How is IPFS different from Cloud?

- Content retrieval based on what the content is, not where the content is
- No dependency on the internet backbone
- It is not structured on ownership and access, but possession and participation



MattOber, *The IPFS Cloud*

Properties of IPFS

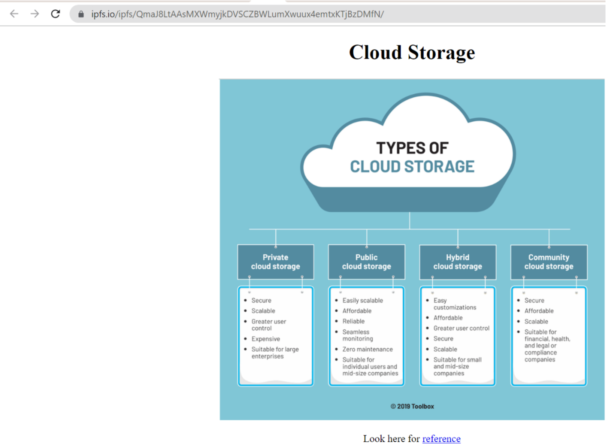
- IPFS is a protocol.
- IPFS is a file system.
- IPFS is a web.
- IPFS is modular.
- IPFS uses crypto.
- IPFS is peer to peer (p2p).
- IPFS is a CDN (Content Delivery Network)
- Does not rely on DNS (Domain Name System) or Certificate Authority System

Decentralization with IPFS

- InterPlanetary File System - Strives to build a system that works across places as disconnected or as far apart as planets.
- Can speed up the web when far away or disconnected.
- Makes it harder to censor content.
- Improves humanity’s access to information
- Persistent data storage
- Uses Merkle DAG (Directed Acyclic Graph) to implement content addressing and fragment downloading of files.

Content Addressing in IPFS

- Types of cloud storage - Click here



ipfs.io/ipfs/QmaJ8LtAAsMXWmyjkDVSCZBWLumXwuux4emtxKTjBzDMfN/

Cloud Storage

Private cloud storage:
- Secure
- Scalable
- Greater user control
- Expensive
- Suitable for large enterprises

Public cloud storage:
- Easily scalable
- Affordable
- Reliable
- Seamless monitoring
- Zero maintenance
- Suitable for individual users and mid-size companies

Hybrid cloud storage:
- Easy customizations
- Affordable
- Greater user control
- Secure
- Scalable
- Suitable for small and mid-size companies

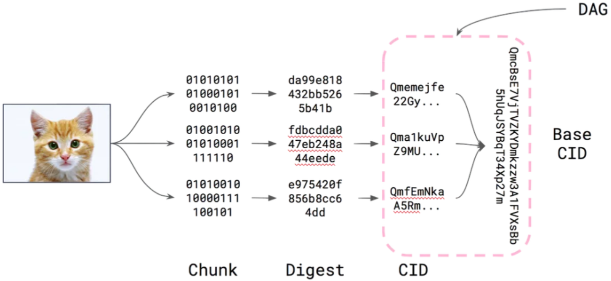
Community cloud storage:
- Secure
- Affordable
- Scalable
- Suitable for financial, health, and legal or compliance companies

© 2019 Toolbox

Look here for reference

Content Addressing in IPFS

- IPFS addresses a file by content. CID (Content Identifier)
- IPFS SHA-25 produces a 256 bit (32 byte) output, encoded with Base58



Farmer, *What's really happening when you add a file to IPFS*

1 Introduction

2 Overview of Cloud Data Storage

3 Decentralised Cloud Storage with IPFS

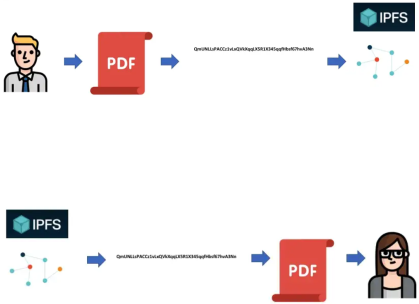
4 Privacy and Security for Cloud Storage

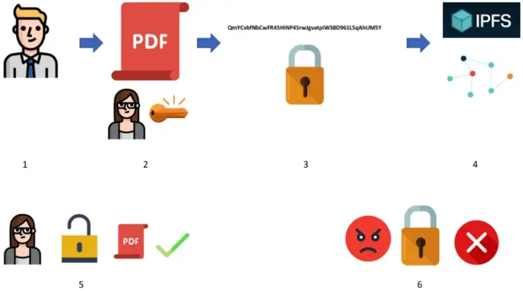
5 Conclusion

Security Threats in Cloud Storage

- Data loss/ leakage
- Data confidentiality
- Accidental exposure of credentials
- Data Location
- Denial of Service
- Lack of trust and dependence on cloud provider
- Misconfigured Cloud Storage
- Cyberattacks

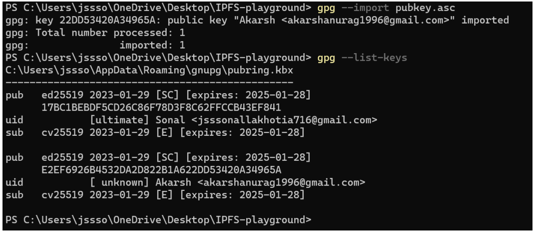
Secure File sharing in IPFS


PDF
QmUNLLsPACCz1vLxQVkXqqLX5R1X345qqfHbsf67hvA3Nn
IPFS
IPFS
QmUNLLsPACCz1vLxQVkXqqLX5R1X345qqfHbsf67hvA3Nn
PDF



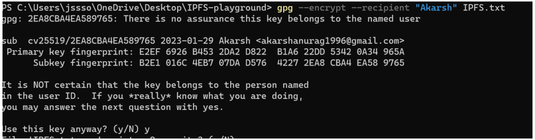
CoralHealth, *Learn to securely share files on the blockchain with IPFS!*

A Simple Demonstration



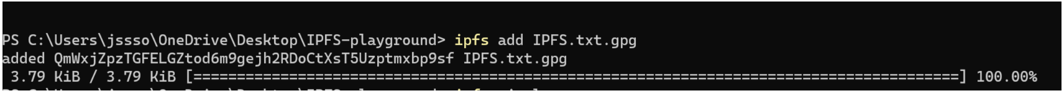
```
PS C:\Users\jssso\OneDrive\Desktop\IPFS-playground> gpg --import pubkey.asc
gpg: key 22DD53420A34965A: public key "Akarsh <akarshanurag1996@gmail.com>" imported
gpg: Total number processed: 1
gpg:               imported: 1
PS C:\Users\jssso\OneDrive\Desktop\IPFS-playground> gpg --list-keys
C:\Users\jssso\AppData\Roaming\gnupg\pubring.kbx
------------------------------------------------
pub   ed25519 2023-01-29 [SC] [expires: 2025-01-28]
      17BC1BEBDF5CD26C86F78D3F8C62FFCCB43EF841
uid           [ultimate] Sonal <jsssonallakhotia716@gmail.com>
sub   cv25519 2023-01-29 [E] [expires: 2025-01-28]

pub   ed25519 2023-01-29 [SC] [expires: 2025-01-28]
      E2EF6926B4532DA2D822B1A622DD53420A34965A
uid           [ unknown] Akarsh <akarshanurag1996@gmail.com>
sub   cv25519 2023-01-29 [E] [expires: 2025-01-28]

PS C:\Users\jssso\OneDrive\Desktop\IPFS-playground>
```

Figure: Adding a public key of a peer to own public keyring



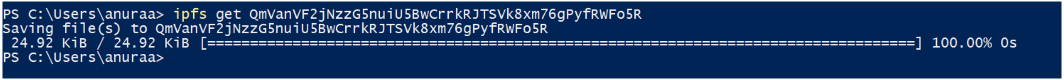
```
PS C:\Users\jssso\OneDrive\Desktop\IPFS-playground> gpg --encrypt --recipient "Akarsh" IPFS.txt
gpg: 2EA8CBA4EA589765: There is no assurance this key belongs to the named user

sub  cv25519/2EA8CBA4EA589765 2023-01-29 Akarsh <akarshanurag1996@gmail.com>
 Primary key fingerprint: E2EF 6926 B453 2DA2 D822  B1A6 22DD 5342 0A34 965A
      Subkey fingerprint: B2E1 016C 4EB7 07DA D576  4227 2EA8 CBA4 EA58 9765

It is NOT certain that the key belongs to the person named
in the user ID.  If you *really* know what you are doing,
you may answer the next question with yes.

Use this key anyway? (y/N) y
```

Figure: Encryping the file using peer public key



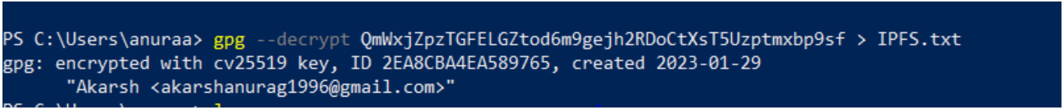
```
PS C:\Users\jssso\OneDrive\Desktop\IPFS-playground> ipfs add IPFS.txt.gpg
added QmWxjZpzTGFELGZtod6m9gejh2RDoCtXsT5Uzptmxbp9sf IPFS.txt.gpg
 3.79 KiB / 3.79 KiB [=====================================================================================] 100.00%
```

Figure: Uploading file to IPFS



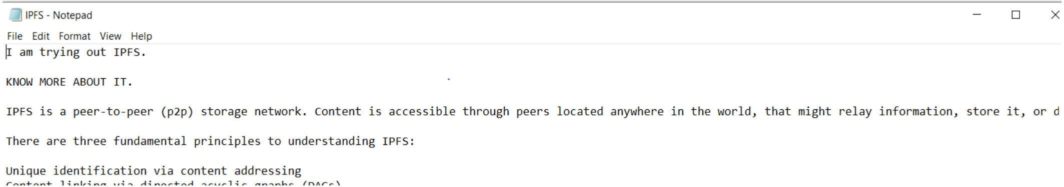
```
PS C:\Users\anuraa> ipfs get QmVanVF2jNzzG5nuiU5BwCrrkRJTSVk8xm76gPyfRWFo5R
Saving file(s) to QmVanVF2jNzzG5nuiU5BwCrrkRJTSVk8xm76gPyfRWFo5R
 24.92 KiB / 24.92 KiB [===================================================================================] 100.00% 0s
PS C:\Users\anuraa>
```

Figure: Downloading the content from IPFS by peer



```
PS C:\Users\anuraa> gpg --decrypt QmWxjZpzTGFELGZtod6m9gejh2RDoCtXsT5Uzptmxbp9sf > IPFS.txt
gpg: encrypted with cv25519 key, ID 2EA8CBA4EA589765, created 2023-01-29
      "Akarsh <akarshanurag1996@gmail.com>"
```

Figure: Decryption by peer



```
I am trying out IPFS.

KNOW MORE ABOUT IT.

IPFS is a peer-to-peer (p2p) storage network. Content is accessible through peers located anywhere in the world, that might relay information, store it, or d

There are three fundamental principles to understanding IPFS:

Unique identification via content addressing
Content linking via directed acyclic graphs (DAGs)
```

Figure: View File

1 Introduction

2 Overview of Cloud Data Storage

3 Decentralised Cloud Storage with IPFS

4 Privacy and Security for Cloud Storage

5 Conclusion

Summary

- Cloud Storage enables scalability, flexibility and security
- It provides sustainability and redundancy.
- It has few disadvantages.
  - Compliance
  - Latency
  - Control
  - Outages
- IPFS isn't a magic cloud that we can freely upload all of our data.
- Data on IPFS is currently being managed similar to centralized database paradigms.
- IPFS supporting infrastructure needs to be built.
- With IPFS a decentralized and secure web is possible

References

AWS. *Types of Cloud Storage*. https://cdcloudlogix.com/different-types-of-aws-cloud-storage/. [Online; accessed 23-02-2023]. 2023.

CoralHealth. *Learn to securely share files on the blockchain with IPFS!* https://mycoralhealth.medium.com/learn-to-securely-share-files-on-the-blockchain-with-ipfs-219ee47df54c. [Online; accessed 23-02-2023].

Farmer, Carson. *What's really happening when you add a file to IPFS*. https://medium.com/textileio/whats-really-happening-when-you-add-a-file-to-ipfs-ae3b8b5e4b0f. [Online; accessed 23-02-2023].

GoogleCloud. *A map of storage options in Google Cloud*. https://cloud.google.com/blog/topics/developers-practitioners/map-storage-options-google-cloud. [Online; accessed 23-02-2023].

MattOber. *The IPFS Cloud*. https://medium.com/pinata/the-ipfs-cloud-352ecaa3ba76. [Online; accessed 23-02-2023].

Wikipedia. *Cloud Storage*. https://en.wikipedia.org/wiki/Cloud_storage. [Online; accessed 23-02-2023].