

# Lecture Six/Seven

Cloud Storage and data cloud

# Overview

Traditional Storage architectures

Cloud storage theory and implementations

Cloud storage and storage in the Cloud.

Cloud storage and Cloud Distributed storage.

**Object storage: swift example**

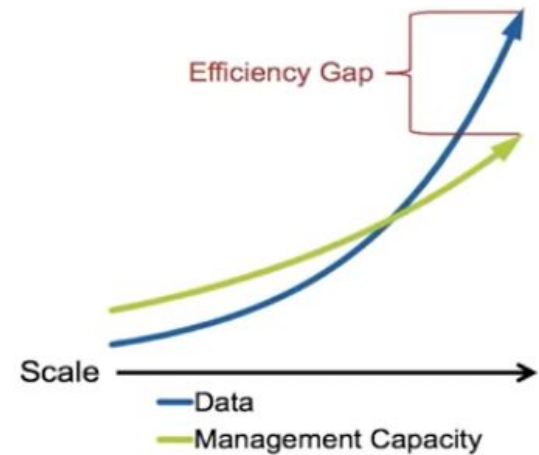
**Data cloud**

# Introduction

*Cloud storage is a service model in which data is maintained, managed and backed up remotely and made available to users over a network (typically the Internet).*

## Data explosion and Mobile device growth

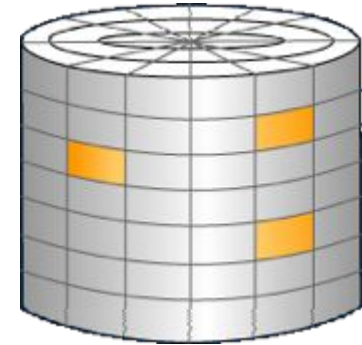
- improve scalability (up+out) and security
- improve performance
- simplify storage management
- on demand access
- unstructured data



# Traditional Storage architecture for a DC

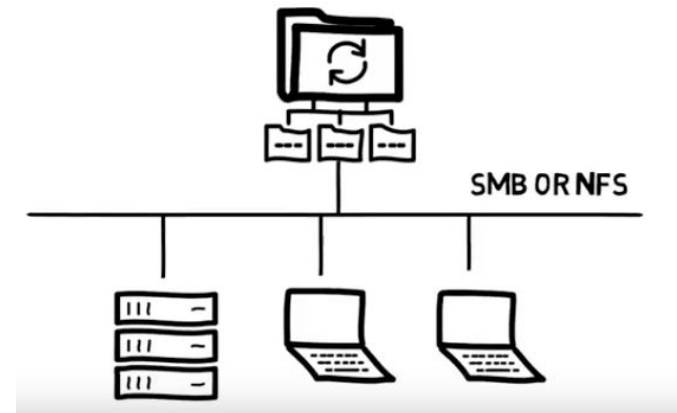
- Block storage

- Volumes
- Blocks (read and Write)
- Fibre Channel or iSCSI protocol
- Local
- Low Latency, high IOPs, low size (<1PB)
- Complex to expand and expensive



- File Storage

- Files and directories
- Network Shared (LAN)
- SMB, NFS, OCFS etc
- High throughput, large size (PBs)
- Scale out capabilities
- Multi-tiered architecture
- Expensive



# Distributed Filesystems: introduction

File system that is **shared** by many distributed clients

The resources (file+dir) on a particular machine are **local** to itself.

Resources on other machines are **remote**

Basic layer for many distributed systems (clients) and applications

A DFS provides a service for clients. The server interface is the normal set of file operations: create, read, etc. on files.

Servers allow clients to perform operations on resources that resides on servers.

# DFS challenges

## Transparency:

Location: a client cannot tell where a file is located

Migration: a file can transparently move to another server

Replication: multiple copies of a file may exist

Concurrency: multiple clients access the same file

## Flexibility

Servers may be added or replaced

Support for multiple file system types

# DFS challenges

## Dependability

Consistency: conflicts with replication & concurrency

Security: users may have different access rights on clients sharing files & network transmission

Fault tolerance: server crash, availability of files

## Performance

Requests may be distributed across servers

Multiple servers allow higher storage capacity

# DFS challenges

## Caching

Reduce network traffic by retaining recently accessed disk blocks in a cache, so that repeated accesses to the same information can be handled locally.

If required data is not already cached, a copy of data is brought from the server to the user.

Perform accesses on the cached copy.

Files are identified with one master copy residing at the server machine,

Copies of (parts of) the file are scattered in different caches.

Cache Consistency Problem -- Keeping the cached copies consistent with the master file



# DFS: client view

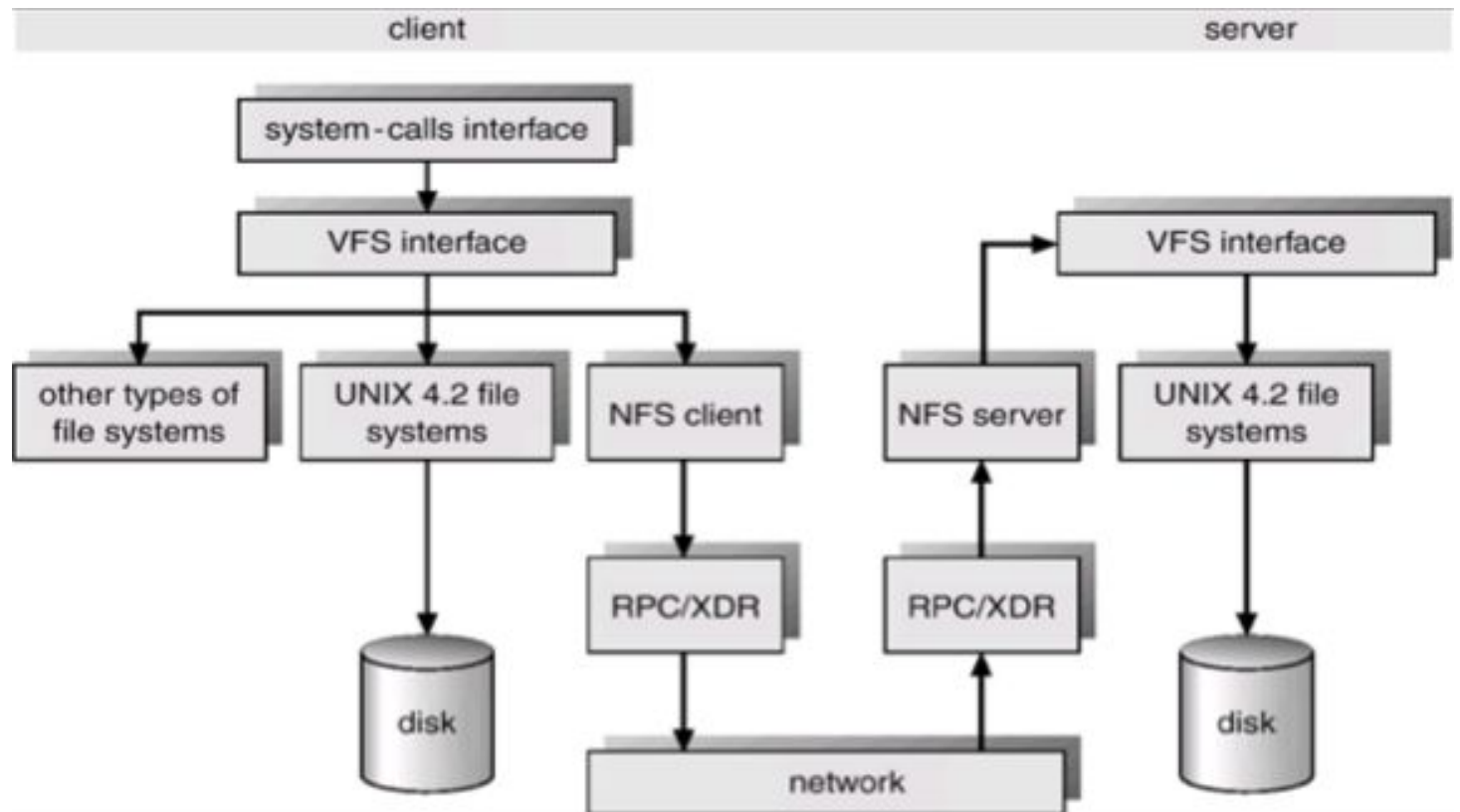
Ideally, the client would perceive remote files like local ones.

Clients, servers, and storage are dispersed across machines.

Configuration and implementation may vary:

Servers may run on dedicated machines, OR  
Servers and clients can be on the same machines.

# DFS: network file system



# Limitation of traditional Storage approach

Handle increasing number of files and users

Growth over geographic and administrative areas

Growth of storage space

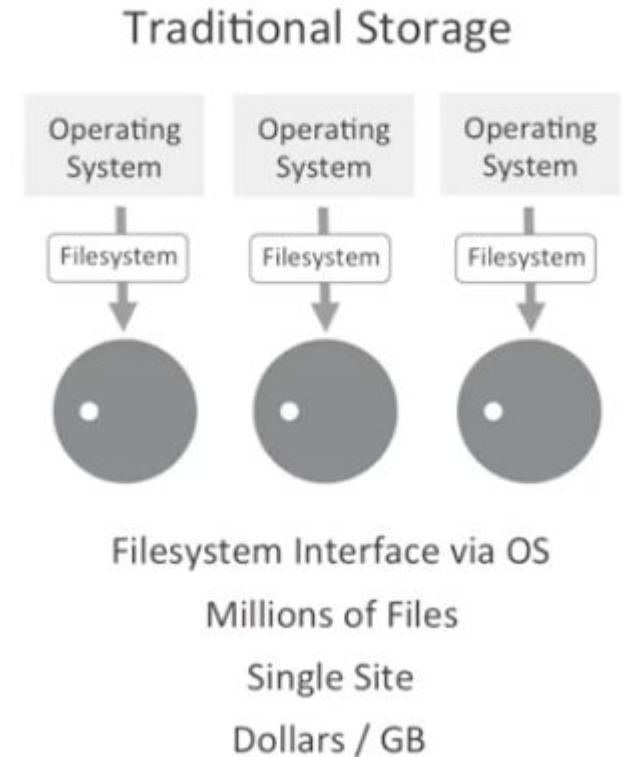
No central naming service

No centralised locking

No central file store

Applications must be aware of volumes/data location

Data is structured and isolated in Filesystems and volumes



# DFS in the cloud: Google File system

## Motivation

- One single distributed file system
- Store big data reliably
- Allow parallel processing of big data

## Assumptions

- Inexpensive components that often fail
- Large files (million of files 100+MB)
- Large streaming reads and small random reads (500Mb/s read/write load)
- Large sequential writes
- Multiple users append to the same file
- High bandwidth is more important than low latency.

# GFS interface

**No common standard like POSIX.**

Provides familiar file system interface:

Create, Delete, Open, Close, Read, Write

*Snapshot*: low cost copy of a whole file with copy-on-write operation

*Record append*: Atomic append operation

# GFS Design Overview

Files split in fixed size chunks of 64 MByte

Chunks stored on chunk servers

Chunks replicated on multiple chunk servers

GFS master manages name space

Clients interact with master to get chunk handles

Clients interact with chunk servers for reads and writes

No explicit caching

# GFS Design

**Master server:** Single master - Keep metadata - accept requests on metadata - Most management activities

**Chunk servers:** Multiple - Keep chunks of data- Accept requests on chunk data

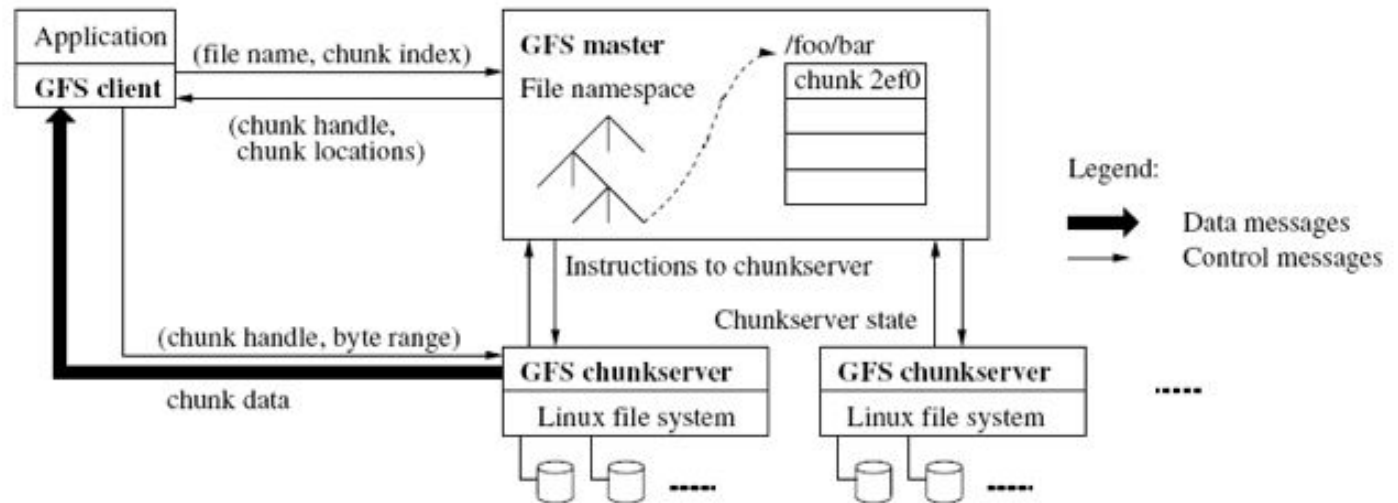


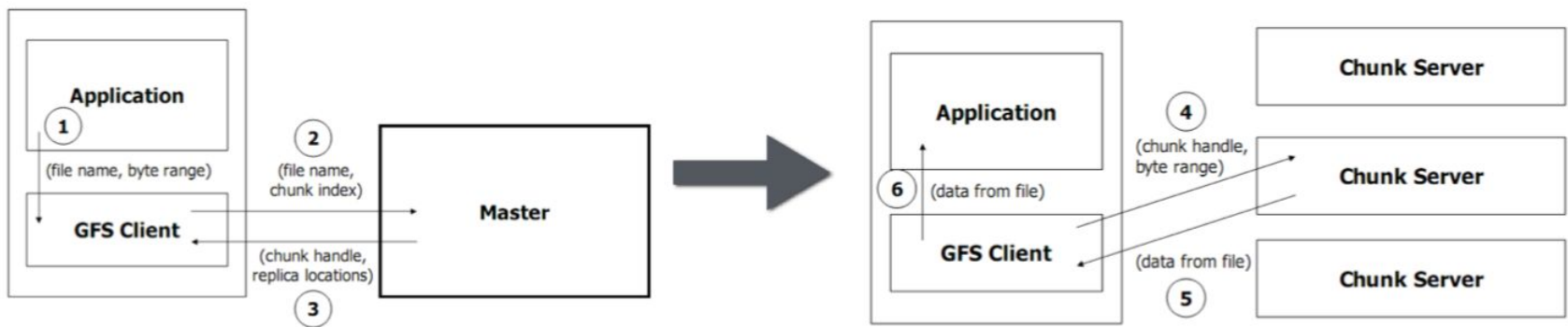
Figure 1: GFS Architecture

# GFS usage

GFS is designed for Google apps and workloads

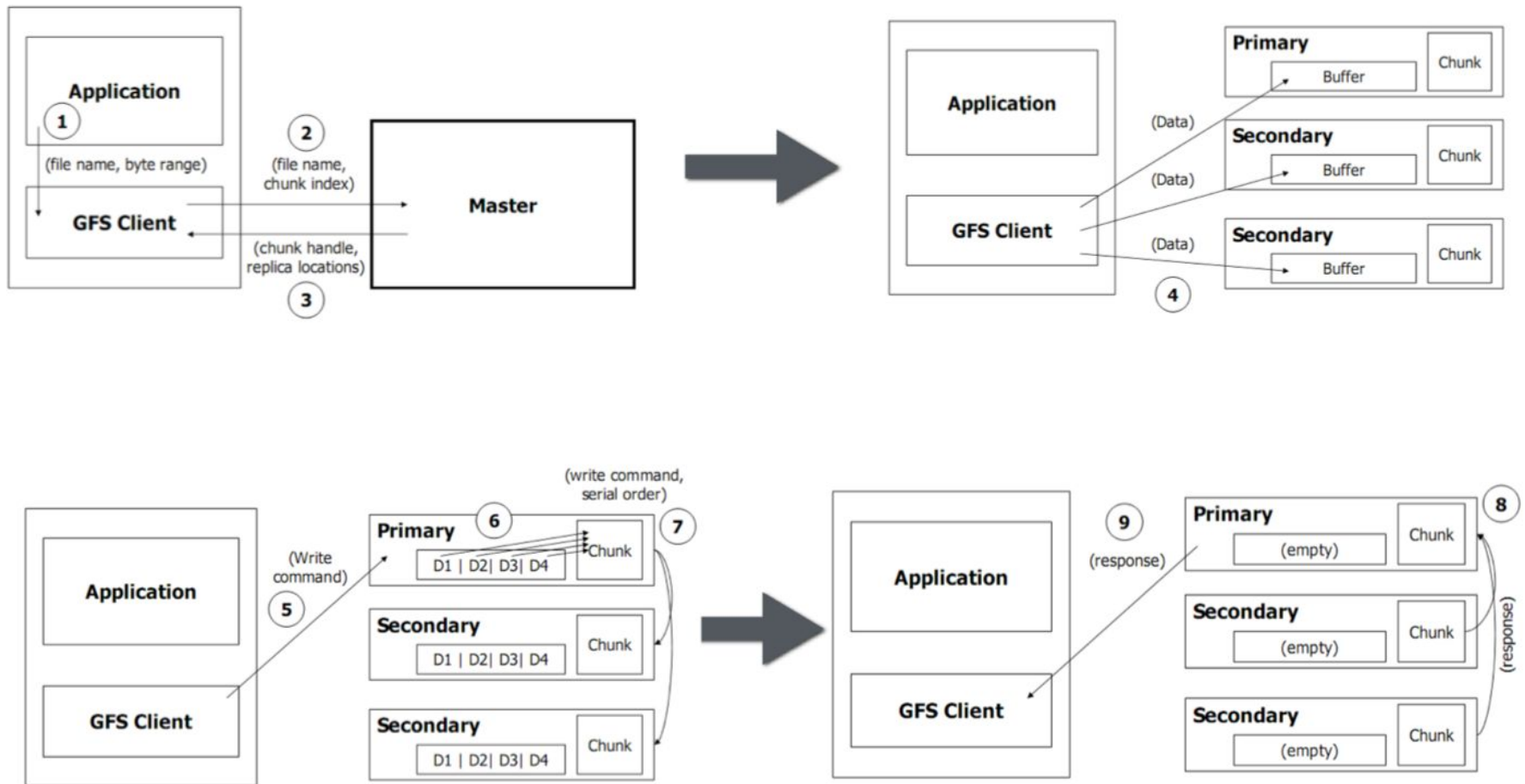
Google apps are designed for GFS

reading operations





# GFS writing operations



# Object storage

Based on Objects (no files/directory)

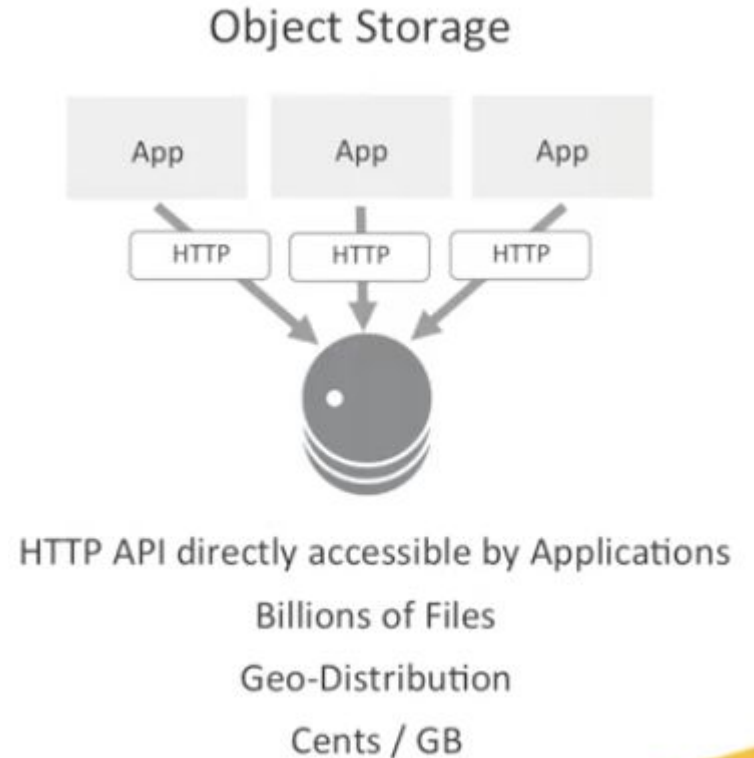
Self descriptive

Linear scale (no hierarchy)

Globally accessible

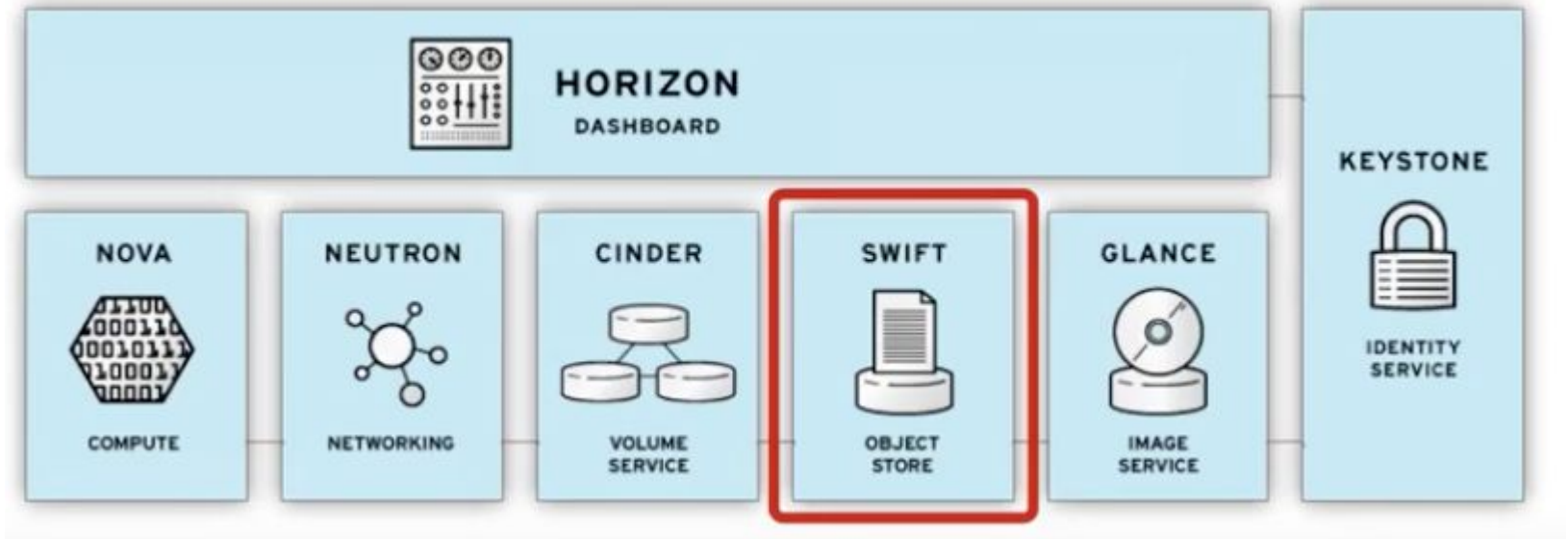
Extensible

Highly parallel

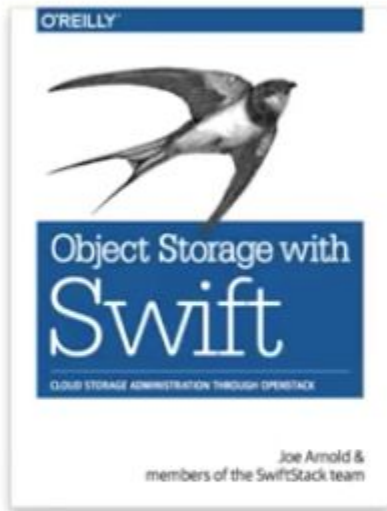


# Example of Object storage

- Public cloud services
  - Amazon S3
  - Google Storage (not Google Drive!!!)
  - HP Cloud Object Storage
  - etc...
- Object Storage Systems
  - OpenStack Object Storage System (swift)



## OpenStack Swift



- OpenStack Swift – Worlds most popular Object Storage System
- Powers extremely large storage clouds
- 100% Open-source (Apache 2)
- Rich ecosystem of tools and applications
- SwiftStack is core developer and leads project
- +180 developers today. Contributors include:



# What is an Object?

An object is a logical unit of storage

- ID (Identification)
- Application data
- Metadata which includes block allocation and length
- Attributes that is accessible by users

Objects have file-like methods

- open, close, read, write

**Object = File + Metadata**

# What is metadata?

## Describes the object

- Helps you to find the right one
- Tells you what it is
- The specifications
- Used where and when
- Access permissions

## Any and all objects

- Different attributes per object
- Add attributes later

Name
utf8Name
type
size
hashScheme
hash
retention
retentionString
retentionClass
changeTimeString
ingestTime
ingestTimeString
accessTime
accessTimeString
updateTime
updateTimeString
uid
gid
permissions
hold
shred
dpl
index
replicated
customMetadata
acl
namespace
objectPath
owner

# Astronomical Data Analysis

## Use Case

Cost effective On-line archiving enriched with metadata.

## Challenge

Astronomical FITS files are self descriptive

Content needs to be available for image reduction

Content must be content of a permanent archival record securely stored

# FITS files

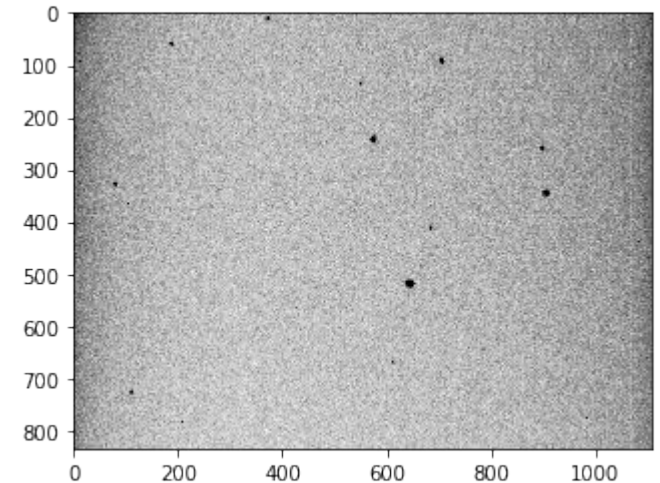
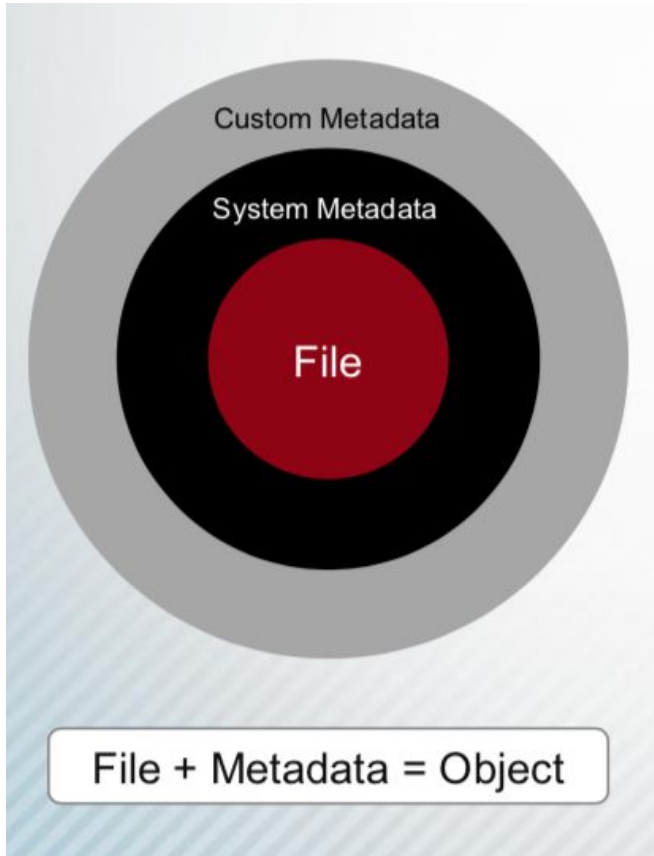
***FITS**. Flexible Image Transport System (**FITS**) is an open standard defining a digital **file format** useful for storage, transmission and processing of data: formatted as *N*-dimensional arrays (for example a 2D image), or tables. **FITS** is the most commonly used digital **file format** in astronomy.*

Data Reduction in python example

[https://github.com/gtaffoni/Learn-Python/blob/master/Lectures/PythonLecture05-Astronomy\\_Data\\_Reduction.ipynb](https://github.com/gtaffoni/Learn-Python/blob/master/Lectures/PythonLecture05-Astronomy_Data_Reduction.ipynb)



# Objects and Metadata



`File class = image`

```
Filename = image.FITS  
Created = Oct 3, 2018 13:29:59  
Last modified: Oct 10, 2018 11:00:00
```

```
OBJECT = 'SZ_Lyn '  
OBSERVER= 'Iafrate '  
TELESCOP= 'SVAS '  
OBJCTRA = '08 09 36'  
OBJCTDEC= '+44 28 18'  
OBJCTALT= ' 84.7258'  
OBJCTAZ = '100.7240'  
OBJCTHA = ' -0.4844'
```

# Mining Metadata

## Metadata:

- stored with the object in extend attributes
- can be changed, removed or added (later)
- can be indexed

Identify sets of related objects based on system and custom metadata

Understand the object store – gather object or content metrics on sets of objects based on metadata

- Content discovery

- Return all files owned by "Bob Smith" created after "3/15/2011" with unexpired retention
- Return objects with a specific retention class defined
- Return objects under retention hold, for mitigation purposes

## Metrics gathering

- What is the size distribution of files in the system?
- What percentage of my files are Word documents?
- Which owners have generated the most content?

# Access an object

Objects have a 64 bit unique ID

Objects live on a flat namespace

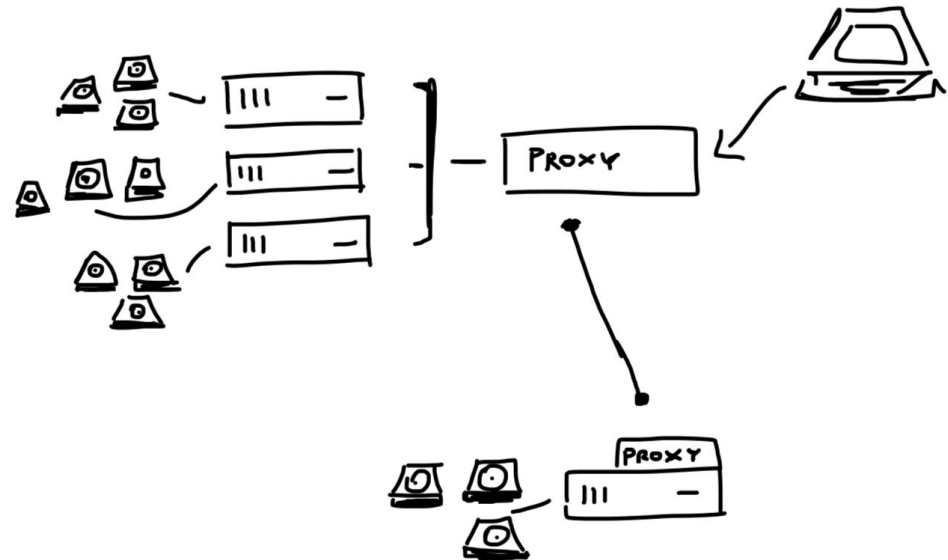
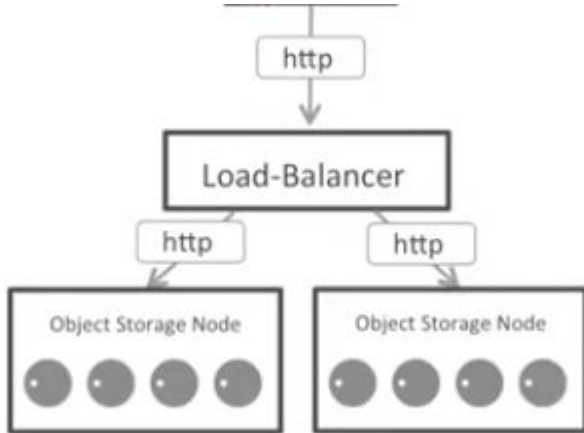
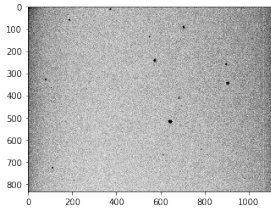
Request for storage are made with HTTP using RESTfull APIs (or SWIFT client)

Three primary components of the request:

1. HTTP Verb (PUT, GET, DELETE, etc.)
2. Authentication Information
3. SURL
4. Metadata (*Optional*)

# HTTP API and SWIFT Architecture

PUT `http://XXX.XXX.XXX.XXX/v1/AUTH_application1/pic/image.fits`



# Manage Objects with HTTP

Base URL      API Version

`http://swift.example.com/v1/AUTH_account/container/object`

- **Storage area** within the cluster
- NOT a user account
- Each account has its own URL
- Swift is multi-tenant

- Namespaces used to group objects within an account
- Containers are unlimited
- Like folders, but can't nest them

- File + Metadata
- Each object is addressed as a URL
- Users name the object
- Objects are not organized based on hierarchy
- Instead, object names may contain "/", so pseudo-nested directories are possible.

Method	Description
GET	Downloads an object with its metadata
PUT	Creates new object with specified data content and metadata
COPY	Copies an object to another object
DELETE	Deletes an object
HEAD	Retrieves object metadata
POST	Creates or updates object metadata

There are also API operations for Accounts and Containers.

# Object Storage Use cases

Video streaming and processing

Map-Reduce

Astronomical data analysis



# SWIFT examples

```
> curl http://X.X.X.X/v1/AUTH_test/cont/obj -X POST -H  
"X-Delete-After: 5" -H "X-Object-Meta-Some: value"
```

```
> swift post -H "X-Delete-After: 5" -m "Some: value"  
(-H "X-Delete-At: 1517210485")
```

```
> curl http://X.X.X.X/v1/AUTH_test/cont/obj -X GET -H  
"X-Auth-Token: AUTH_tk5917..." -H "Range: bytes=0-5"
```

```
( -H "Range: bytes=0-5,16-" )
```

# Swift examples

```
[root@cloud SWIFT(keystone_taffoni)]# swift auth
```

```
export OS_STORAGE_URL=http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16
```

```
export OS_AUTH_TOKEN=b45d1bc7ec69470fa8d75a51a8020c29
```

```
[root@cloud SWIFT(keystone_taffoni)]# swift upload TestFile tests.sh
```

```
tests.sh
```

```
[root@cloud SWIFT(keystone_taffoni)]# swift list
```

```
CADC
```

```
TestFile
```



# Using an Object Storage

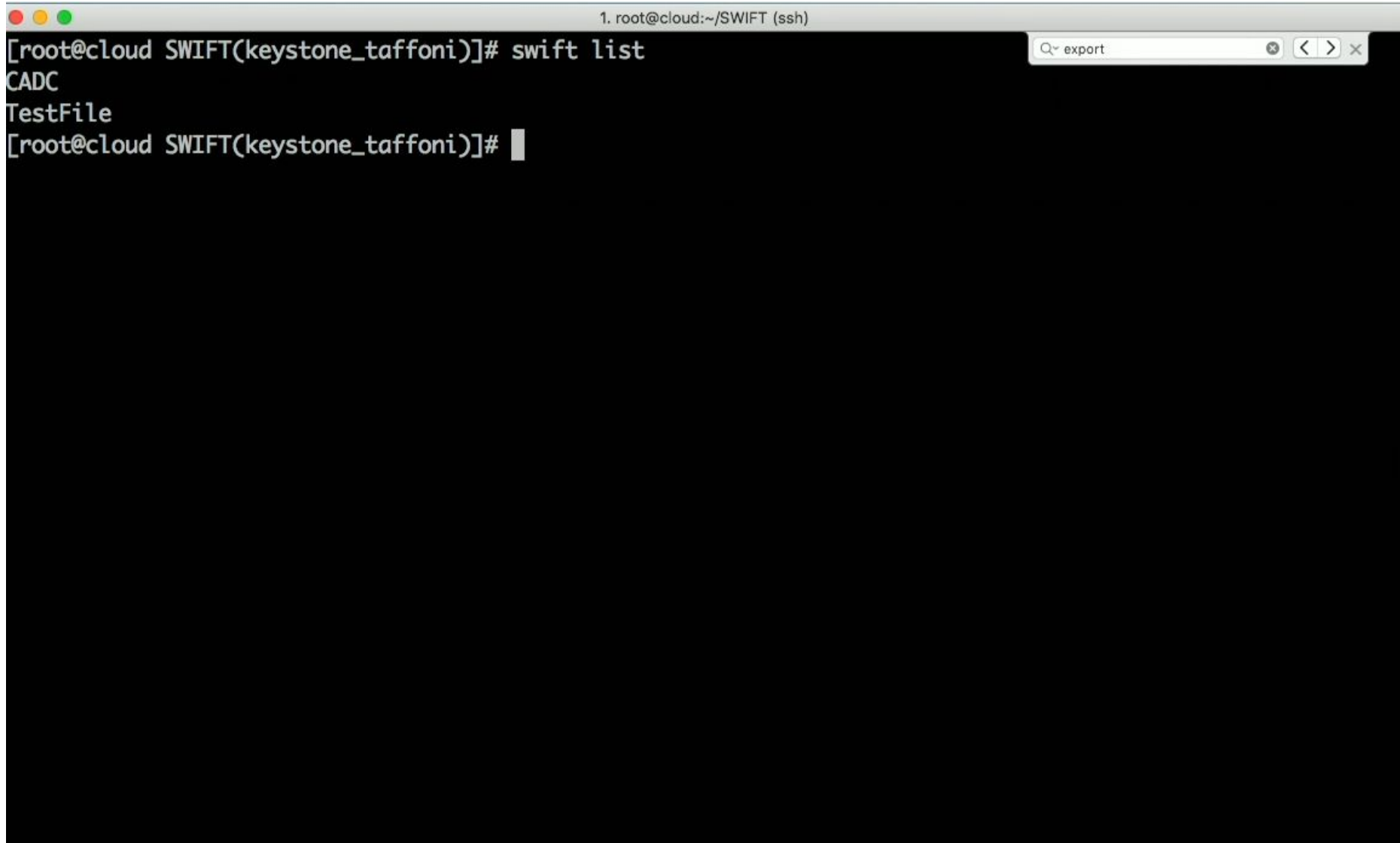
```
import swiftclient
user = 'account_name:username'
key = 'your_api_key'
container_name= 'astro_data'

conn = swiftclient.Connection(
    user=user,
    key=key,
    authurl='https://XXX.XXX.XXX.XXX/auth',
```

```
for data in conn.get_container(container_name)[1]:
    print '{0}\t{1}\t{2}'.format(data['name'], data['bytes'],
data['last_modified'], data['OBSERVER'])
```

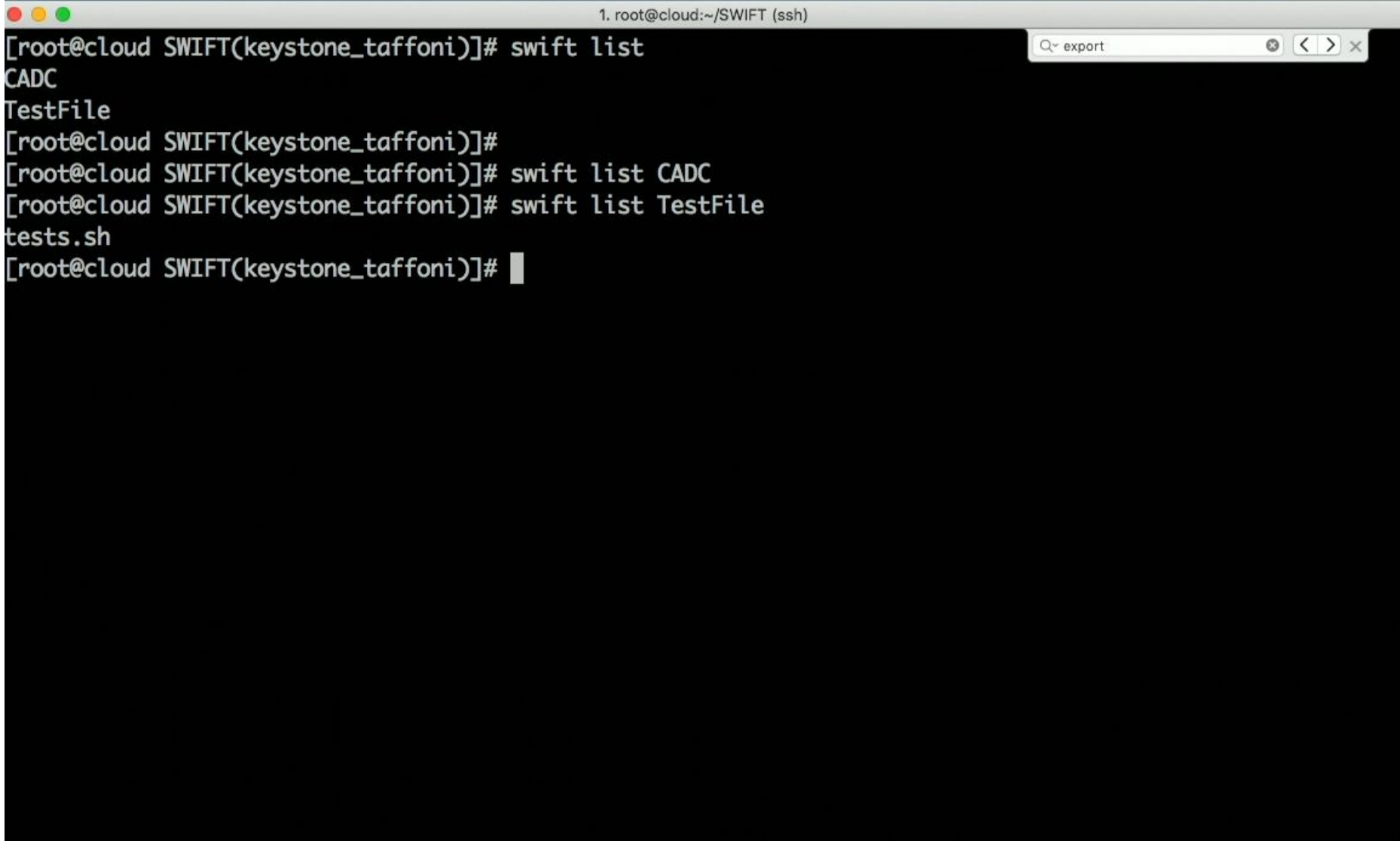
```
obj_tuple = conn.get_object(container_name, 'IMAGE.FITS')
```

# Using an Object Storage

A terminal window with a title bar that reads "1. root@cloud:~/SWIFT (ssh)". The terminal content shows the command "swift list" being executed, resulting in two lines of output: "CADC" and "TestFile". The prompt "[root@cloud SWIFT(keystone\_taffoni)]#" is visible at the beginning and end of the command sequence. A search bar in the top right corner of the terminal window contains the text "export".

```
1. root@cloud:~/SWIFT (ssh)
[root@cloud SWIFT(keystone_taffoni)]# swift list
CADC
TestFile
[root@cloud SWIFT(keystone_taffoni)]#
```

# Using an Object Storage



A terminal window titled "1. root@cloud:~/SWIFT (ssh)" showing the execution of Swift commands. The terminal output is as follows:

```
[root@cloud SWIFT(keystone_taffoni)]# swift list
CADC
TestFile
[root@cloud SWIFT(keystone_taffoni)]#
[root@cloud SWIFT(keystone_taffoni)]# swift list CADC
[root@cloud SWIFT(keystone_taffoni)]# swift list TestFile
tests.sh
[root@cloud SWIFT(keystone_taffoni)]#
```

# Using an Object Storage

```
1. root@cloud:~/SWIFT (ssh)
[root@cloud SWIFT(keystone_taffoni)]# swift list
CADC
TestFile
[root@cloud SWIFT(keystone_taffoni)]#
[root@cloud SWIFT(keystone_taffoni)]# swift list CADC
[root@cloud SWIFT(keystone_taffoni)]# swift list TestFile
tests.sh
[root@cloud SWIFT(keystone_taffoni)]# swift auth
export OS_STORAGE_URL=http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16
export OS_AUTH_TOKEN=ab1a0778fad94ba1ab708bc20a5ecf07
[root@cloud SWIFT(keystone_taffoni)]#
```

# Using an Object Storage

```
1. root@cloud:~/SWIFT (ssh)
HTTP/1.1 201 Created
Content-Length: 0
Content-Type: text/html; charset=UTF-8
X-Trans-Id: tx157ab65ad2754660aeebf-005bfd79db
Date: Tue, 27 Nov 2018 17:07:39 GMT

[root@cloud SWIFT(keystone_taffoni)]# swift list
CADC
TestFile
odmec

[root@cloud SWIFT(keystone_taffoni)]# swift list odmec
[root@cloud SWIFT(keystone_taffoni)]# curl -i -T tests.sh -X PUT -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07" http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/odmec/myobject
HTTP/1.1 100 Continue

HTTP/1.1 201 Created
Last-Modified: Tue, 27 Nov 2018 17:09:22 GMT
Content-Length: 0
Etag: 58d6883581fcd9abc05776d8ed73a33c
Content-Type: text/html; charset=UTF-8
X-Trans-Id: txb55411f145d14008baa55-005bfd7a41
Date: Tue, 27 Nov 2018 17:09:21 GMT

[root@cloud SWIFT(keystone_taffoni)]# swift list odmec
```

# Using an Object Storage

```
1. root@cloud:~/SWIFT (ssh)
env[ST_AUTH_VERSION], env[OS_AUTH_VERSION],
env[OS_IDENTITY_API_VERSION] or 1.0.
-U USER, --user=USER User name for obtaining an auth token.
-K KEY, --key=KEY Key for obtaining an auth token.
-R RETRIES, --retries=RETRIES
    The number of times to retry a failed connection.
--insecure Allow swiftclient to access servers without having to
    verify the SSL certificate. Defaults to
    env[SWIFTCLIENT_INSECURE] (set to 'true' to enable).
--no-ssl-compression This option is deprecated and not used anymore. SSL
    compression should be disabled by default by the
    system SSL library.
[root@cloud SWIFT(keystone_taffoni)]# curl -i --head -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07"
http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/odmec/myobject
HTTP/1.1 200 OK
Content-Length: 280
Accept-Ranges: bytes
Last-Modified: Tue, 27 Nov 2018 17:09:22 GMT
Etag: 58d6883581fcd9abc05776d8ed73a33c
X-Timestamp: 1543338561.34128
Content-Type: application/octet-stream
X-Trans-Id: tx84fc6715e77245cc8d0a5-005bfd7acf
Date: Tue, 27 Nov 2018 17:11:43 GMT
[root@cloud SWIFT(keystone_taffoni)]#
```



# Using an Object Storage

```
1. root@cloud:~/SWIFT (ssh) 🔔
verify the SSL certificate. Defaults to
env[SWIFTCLIENT_INSECURE] (set to 'true' to enable).
--no-ssl-compression This option is deprecated and not used anymore. SSL
compression should be disabled by default by the
system SSL library.
[root@cloud SWIFT(keystone_taffoni)]# curl -i --head -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07"
http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/odmec/myobject
HTTP/1.1 200 OK
Content-Length: 280
Accept-Ranges: bytes
Last-Modified: Tue, 27 Nov 2018 17:09:22 GMT
Etag: 58d6883581fcd9abc05776d8ed73a33c
X-Timestamp: 1543338561.34128
Content-Type: application/octet-stream
X-Trans-Id: tx84fc6715e77245cc8d0a5-005bfd7acf
Date: Tue, 27 Nov 2018 17:11:43 GMT
[root@cloud SWIFT(keystone_taffoni)]# curl -s -S -X DELETE -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5e
cf07" http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/odmec/myobject
[root@cloud SWIFT(keystone_taffoni)]# swift list
CADC
TestFile
odmec
[root@cloud SWIFT(keystone_taffoni)]# swift list odmec
[root@cloud SWIFT(keystone_taffoni)]# █
```

```
1. root@cloud:~/SWIFT (ssh)
Content-Length: 0
X-Object-Meta-Mtime: 1543329870.786404
X-Copied-From-Last-Modified: Tue, 27 Nov 2018 14:43:45 GMT
X-Copied-From: TestFile/tests.sh
Last-Modified: Tue, 27 Nov 2018 17:18:21 GMT
Etag: 58d6883581fcd9abc05776d8ed73a33c
X-Copied-From-Account: AUTH_ff8fcc76cdb14e1a8568202b89d34a16
Content-Type: text/html; charset=UTF-8
X-Trans-Id: txbd99545211b14898b30e8-005bfd7c5c
Date: Tue, 27 Nov 2018 17:18:20 GMT

[root@cloud SWIFT(keystone_taffoni)]# swift list odmec
myobj]
[root@cloud SWIFT(keystone_taffoni)]# curl -i -X PUT -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07
-H "X-Object-Meta-Mio: pippo" http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/TestF
e/tests.sh
HTTP/1.1 411 Length Required
Content-Length: 30
Content-Type: text/plain
X-Trans-Id: txacbe97663e5d467fb8734-005bfd7cbc
Date: Tue, 27 Nov 2018 17:19:56 GMT
```



# Using an Object Storage

```
1. root@cloud:~/SWIFT (ssh)
HTTP/1.1 204 No Content
Content-Length: 0
Content-Type: text/html; charset=UTF-8
X-Trans-Id: tx2839d94436864730bd007-005bfd7d47
Date: Tue, 27 Nov 2018 17:22:15 GMT

[root@cloud SWIFT(keystone_taffoni)]# curl -i --head-H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07" h
http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/TestFile
curl: option --head-H: is unknown
curl: try 'curl --help' or 'curl --manual' for more information

[root@cloud SWIFT(keystone_taffoni)]# curl -i --head -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07"
http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/TestFile
HTTP/1.1 204 No Content
Content-Length: 0
X-Container-Object-Count: 1
X-Container-Meta-Mymeta: pippo
Accept-Ranges: bytes
X-Storage-Policy: Policy-0
X-Container-Bytes-Used: 280
X-Timestamp: 1543329824.44274
Content-Type: text/plain; charset=utf-8
X-Trans-Id: tx8412cc81f4b04c79a6e47-005bfd7d74
Date: Tue, 27 Nov 2018 17:23:00 GMT

[root@cloud SWIFT(keystone_taffoni)]#
```

# Using Swift: print file content

```
1. root@cloud:~/SWIFT (ssh)
Content-Length: 0
Content-Type: text/html; charset=UTF-8
X-Trans-Id: tx2839d94436864730bd007-005bfd7d47
Date: Tue, 27 Nov 2018 17:22:15 GMT

[root@cloud SWIFT(keystone_taffoni)]# curl -i --head-H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07" http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/TestFile
curl: option --head-H: is unknown
curl: try 'curl --help' or 'curl --manual' for more information
[root@cloud SWIFT(keystone_taffoni)]# curl -i --head -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07" http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/TestFile
HTTP/1.1 204 No Content
Content-Length: 0
X-Container-Object-Count: 1
X-Container-Meta-Mymeta: pippo
Accept-Ranges: bytes
X-Storage-Policy: Policy-0
X-Container-Bytes-Used: 280
X-Timestamp: 1543329824.44274
Content-Type: text/plain; charset=utf-8
X-Trans-Id: tx8412cc81f4b04c79a6e47-005bfd7d74
Date: Tue, 27 Nov 2018 17:23:00 GMT

[root@cloud SWIFT(keystone_taffoni)]# curl -i -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf07" http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/TestFile/tests.sh
```

# Using Swift: file content

```
1. root@cloud:~/SWIFT (ssh)
140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/TestFile/tests.s
HTTP/1.1 200 OK
Content-Length: 280
Content-Type: application/x-sh
Accept-Ranges: bytes
Last-Modified: Tue, 27 Nov 2018 14:43:45 GMT
Etag: 58d6883581fcd9abc05776d8ed73a33c
X-Timestamp: 1543329824.79508
X-Object-Meta-Mtime: 1543329870.786404
X-Trans-Id: tx1f3c50547428408e8d796-005bfd7da5
Date: Tue, 27 Nov 2018 17:23:49 GMT

curl -i \
  -H "Content-Type: application/json" \
  -d '{
    "auth": {
      "passwordCredentials": {
        "password": "-----",
        "username": "taffoni"
      },
      "tenantName": "OATs"
    }
  }' "https://cloud.oats.inaf.it:5000/v2.0/tokens" ; echo
[root@cloud SWIFT(keystone_taffoni)]#
```



# Using Swift: range

```
1. root@cloud:~/SWIFT (ssh)
Accept-Ranges: bytes
Last-Modified: Tue, 27 Nov 2018 14:43:45 GMT
Content-Range: bytes 0-5/280
Etag: 58d6883581fcd9abc05776d8ed73a33c
X-Timestamp: 1543329824.79508
X-Object-Meta-Mtime: 1543329870.786404
X-Trans-Id: tx137e9d6453eb4e01a15e6-005bfd7df3
Date: Tue, 27 Nov 2018 17:25:07 GMT

curl -[root@cloud SWIFT(keystone_taffo^C
[root@cloud SWIFT(keystone_taffoni)]# curl -i -r 10-15 -H "X-Auth-Token: ab1a0778fad94ba1ab708bc20a5ecf
07" http://140.105.74.223:8080/v1/AUTH_ff8fcc76cdb14e1a8568202b89d34a16/TestFile/tests.sh
HTTP/1.1 206 Partial Content
Content-Length: 6
Content-Type: application/x-sh
Accept-Ranges: bytes
Last-Modified: Tue, 27 Nov 2018 14:43:45 GMT
Content-Range: bytes 10-15/280
Etag: 58d6883581fcd9abc05776d8ed73a33c
X-Timestamp: 1543329824.79508
X-Object-Meta-Mtime: 1543329870.786404
X-Trans-Id: txc831b1ffb2cd45b488d82-005bfd7e1d
Date: Tue, 27 Nov 2018 17:25:49 GMT

-H "[root@cloud SWIFT(keystone_taffoni)]# █
```

# AMAZON S3 object storage

- Amazon Simple Storage Service is an object storage.
- Amazon S3 has a simple web services interface that you can use to store and retrieve any amount of data, at any time, from anywhere on the web.
- It provides access to developers through REST APIs or proprietary APIs.
- It is based on buckets and objects
- Files from 1 byte to 5 TBs

# AMAZON S3 object storage

The screenshot shows the AWS Management Console interface. At the top, there's a navigation bar with the AWS logo, a menu for 'Services', 'Resource Groups', and specific services like 'Route 53', 'EC2', and 'Simple Email Service'. The main content area is divided into several sections:

- AWS services:** A search bar with the placeholder text 'Find a service by name or feature (for example, EC2, S3 or VM, storage)'. Below it, a 'Recently visited services' section lists 'Systems Manager', 'Billing', 'RDS', 'EC2', and 'Cost Explorer'. A link for 'All services' is also present.
- Build a solution:** A section titled 'Get started with simple wizards and automated workflows.' containing six solution cards:
  - Launch a virtual machine:** With EC2, ~2-3 minutes.
  - Build a web app:** With Elastic Beanstalk, ~6 minutes.
  - Build using virtual servers:** With Lightsail, ~1-2 minutes.
  - Connect an IoT device:** With AWS IoT, ~5 minutes.
  - Start a development project:** With CodeStar, ~5 minutes.
  - Register a domain:** With Route 53, ~3 minutes.A 'See more' link is at the bottom of this section.
- Helpful tips:** Two tips are listed:
  - Manage your costs:** Get real-time billing alerts based on your cost and usage budgets. [Start now](#)
  - Create an organization:** Use AWS Organizations for policy-based management of multiple AWS accounts. [Start now](#)
- Explore AWS:** A section for 'Amazon Relational Database Service (RDS)' with a description: 'RDS manages and scales your database for you. RDS supports Aurora, MySQL, PostgreSQL, MariaDB, Oracle, and SQL Server. [Learn more.](#)' Below this is a section for 'Real-Time Analytics with Amazon Kinesis' with a description: 'Stream and analyze real-time data, so you can get timely insights and react quickly. [Learn more.](#)'

# AMAZON S3 object storage

The screenshot displays the AWS Management Console interface for creating a new S3 bucket. A modal window titled "Create bucket" is open, showing a four-step process: 1. Name and region, 2. Set properties, 3. Set permissions, and 4. Review. The current step is "Name and region".

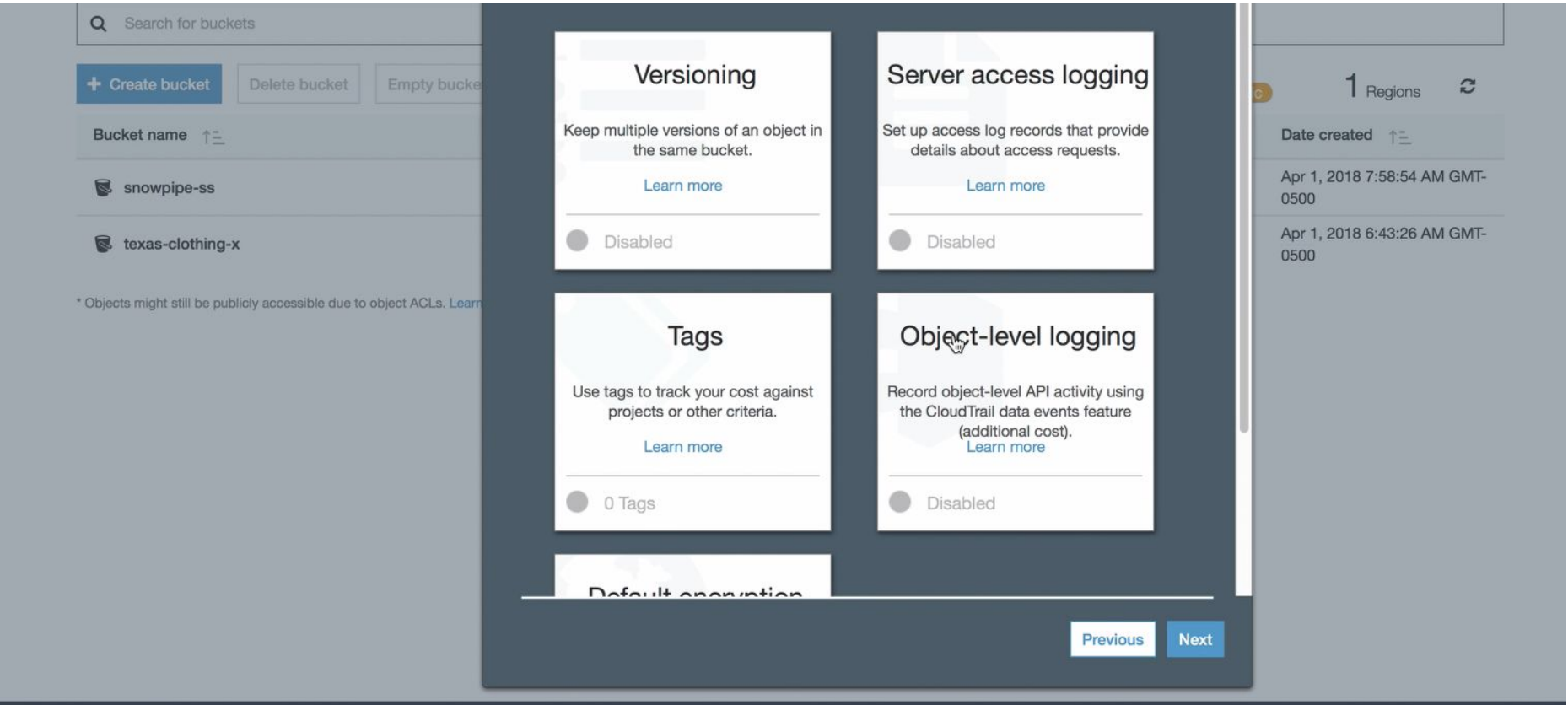
Key elements in the "Name and region" step include:

- Bucket name:** A text input field with a placeholder "Enter DNS-compliant bucket name".
- Region:** A dropdown menu currently set to "US East (N. Virginia)".
- Copy settings from an existing bucket:** A section with a dropdown menu labeled "Select bucket (optional)" and "2 Buckets" shown.

At the bottom of the modal, there are three buttons: "Create", "Cancel", and "Next".

The background shows the AWS S3 console with a search bar, a "Create bucket" button, and a list of buckets including "snowpipe-ss" and "texas-clothing-x".

# AMAZON S3 object storage





# AMAZON S3 object storage

The screenshot displays the Amazon S3 console interface. A modal window titled "Versioning" is open, allowing configuration for a bucket. The modal contains three main sections:

- Versioning:** Features two radio button options. "Enable versioning" is selected, and "Suspend versioning" is unselected. A description for "Suspend versioning" states: "This suspends the creation of object versions for all operations but preserves any existing object versions." Buttons for "Cancel" and "Save" are located at the bottom right of this section.
- Server access logging:** Includes a description: "Set up access log records that provide details about access requests." and a "Learn more" link. A toggle switch is currently set to "Disabled".
- Tags:** Includes a description: "Use tags to track your cost against projects or other criteria." and a "Learn more" link. A toggle switch is currently set to "0 Tags".

At the bottom of the modal, there are navigation buttons: "Previous" and "Next". A "Pausa (k)" button is also visible at the bottom left of the modal area. The background shows a bucket list with "snowpipe-ss" and "texas-clothing-x".

# AMAZON S3 object storage

[Upload](#) [+ Create folder](#) [More](#) ▾

US East (N. Virginia) 

This bucket is empty. Upload new objects to get started.



## Upload an object

Buckets are globally unique containers for everything that you store in Amazon S3.

[Learn more](#)



## Set object properties

After you create a bucket, you can upload your objects (for example, your photo or video files).

[Learn more](#)



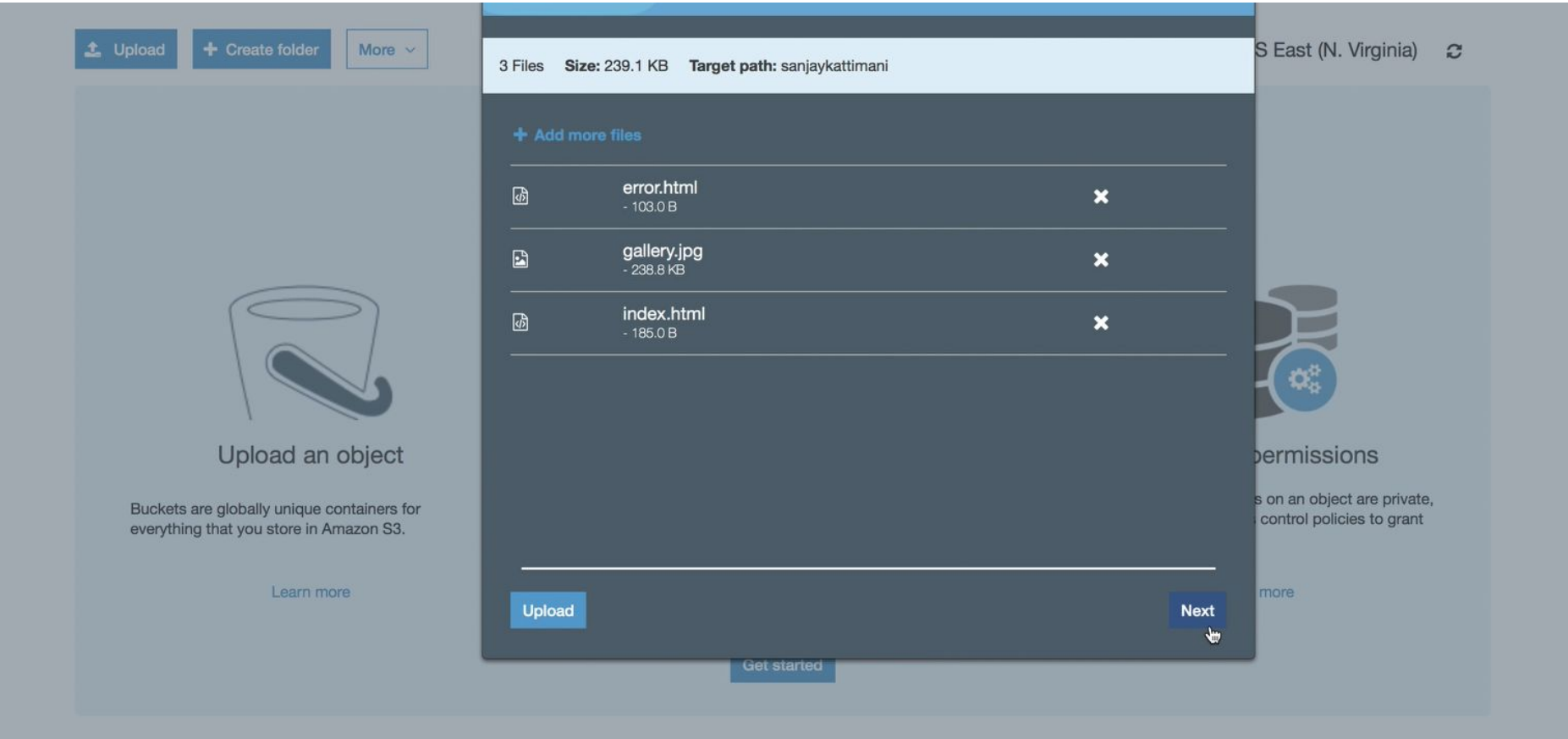
## Set object permissions

By default, the permissions on an object are private, but you can set up access control policies to grant permissions to others.

[Learn more](#)

[Get started](#)

# AMAZON S3 object storage



# AMAZON S3 object storage

The screenshot shows the Amazon S3 console interface in a Safari browser. The address bar displays the URL: `s3.console.aws.amazon.com/s3/object/sanjaykattimani/index.html?region=us-east-`. The navigation bar includes the AWS logo, 'Services', 'Resource Groups', and icons for 'Route 53', 'EC2', and 'Simple Email Service'. The user's name 'kattimani' and 'Global' region are also visible.

The main content area shows the object 'index.html' with a 'Latest version' dropdown. Below this are four tabs: 'Overview', 'Properties', 'Permissions', and 'Select from'. The 'Properties' tab is active, displaying several action buttons: 'Open', 'Download', 'Download as', 'Make public', and 'Copy path'. A mouse cursor is hovering over the 'Make public' button.

Below the buttons, the following object properties are listed:

- Owner:** sanjaycov
- Last modified:** Apr 11, 2018 8:56:46 PM GMT-0500
- Etag:** eff860d85de4e21cc73019d3e2790e70
- Storage class:** Standard
- Server-side encryption:** None
- Size:** 185
- Link:** <https://s3.amazonaws.com/sanjaykattimani/index.html>

# AMAZON S3 object storage

The screenshot shows the Amazon S3 console interface in a Safari browser window. The address bar displays the URL: `s3.console.aws.amazon.com/s3/buckets/sanjaykattimani/?region=us-east-1&tab=pr`. The main content area features several configuration panels:

- Static website hosting:** A modal dialog with a blue header and a close button. It shows the endpoint `http://sanjaykattimani.s3-website-us-east-1.amazonaws.com`. Three radio buttons are present: "Use this bucket to host a website" (unselected), "Redirect requests" (unselected), and "Disable website hosting" (selected). "Cancel" and "Save" buttons are at the bottom.
- Object-level logging:** A panel with a grey header. It describes recording API activity using CloudTrail data events (additional cost) and includes a "Learn more" link. A toggle switch is currently "Disabled".
- Default encryption:** A panel with a grey header. It describes automatically encrypting objects stored in Amazon S3 and includes a "Learn more" link. A toggle switch is currently "Disabled".

Below these panels is the "Advanced settings" section, which includes three cards: "Tags" (Use tags to track your cost against), "Transfer acceleration" (Enable fast, easy and secure transfers of), and "Events" (Receive notifications when specific).

# AMAZON S3 object storage

The screenshot shows the Amazon S3 console interface in a Safari browser window. The browser's address bar displays the URL: `s3.console.aws.amazon.com/s3/buckets/sanjaykattimani/?region=us-east-1&tab=pr`. The main content area is divided into two panels. The left panel, titled "Static website hosting", contains the following configuration options: 

- Endpoint: `http://sanjaykattimani.s3-website-us-east-1.amazonaws.com`
- Radio button selected: "Use this bucket to host a website" (with a "Learn more" link)
- Index document: `in`
- Error document: `error.html`
- Redirection rules (optional): An empty text area.
- Radio buttons: "Redirect requests" (unselected) and "Disable website hosting" (unselected).
- Buttons: "Cancel" and "Save".

The right panel, titled "Object-level logging", contains the following information: 

- Text: "Record object-level API activity using the CloudTrail data events feature (additional cost). (with a "Learn more" link)"
- Radio button: "Disabled" (unselected).

At the bottom left, a partially visible panel titled "Default encryption" is shown.

# AMAZON S3 object storage

The screenshot shows the Amazon S3 console interface in a Safari browser window. The address bar indicates the URL: `s3.console.aws.amazon.com/s3/buckets/sanjaykattimani/?region=us-east-1&tab=pr`. The main content area displays several configuration cards for the bucket 'sanjaykattimani':

- Versioning:** A card with the text 'Keep multiple versions of an object in the same bucket.' and a 'Learn more' link. The status is 'Disabled'.
- Static website hosting:** A card with the text 'Host a static website, which does not require server-side technologies.' and a 'Learn more' link. The status is 'Bucket hosting' (checked).
- Object-level logging:** A card with the text 'Record object-level API activity using the CloudTrail data events feature (additional cost).' and a 'Learn more' link. The status is 'Disabled'.
- Default encryption:** A card with the text 'Automatically encrypt objects when stored in Amazon S3' and a 'Learn more' link. The status is 'Disabled'.

An 'Advanced settings' section is visible at the bottom of the page. A modal window titled 'Server access logging' is open, showing the following configuration options:

- Enable logging
- Target bucket: `sanjaykattimani` (dropdown menu)
- Target prefix: `Enter target prefix` (input field with an information icon)
- Disable logging
- Buttons: 'Cancel' and 'Save'

# AMAZON S3 object storage

The screenshot shows the Amazon S3 console interface in a Safari browser window. The browser's address bar displays the URL: `s3.console.aws.amazon.com/s3/buckets/sanjaykattimani/?region=us-east-1&tab=pr`. The main content area features four configuration cards: 'Tags', 'Events', 'Requester pays', and 'Transfer acceleration'. The 'Transfer acceleration' dialog box is open, showing the endpoint `sanjaykattimani.s3-accelerate.amazonaws.com` with a red circle around the domain part. Below the endpoint, there is a description: 'Use the new accelerated endpoint for faster data transfers, which will incur an additional fee.' A link to 'Want to compare your data transfer speed by region?' is present. Two radio buttons are shown: 'Enabled' (selected) and 'Suspended'. At the bottom of the dialog are 'Cancel' and 'Save' buttons. The other cards show '0 Tags', '0 Active notifications', and 'Disabled' for 'Requester pays'.

**Tags**  
Use tags to track your cost against projects or other criteria.  
[Learn more](#)  
0 Tags

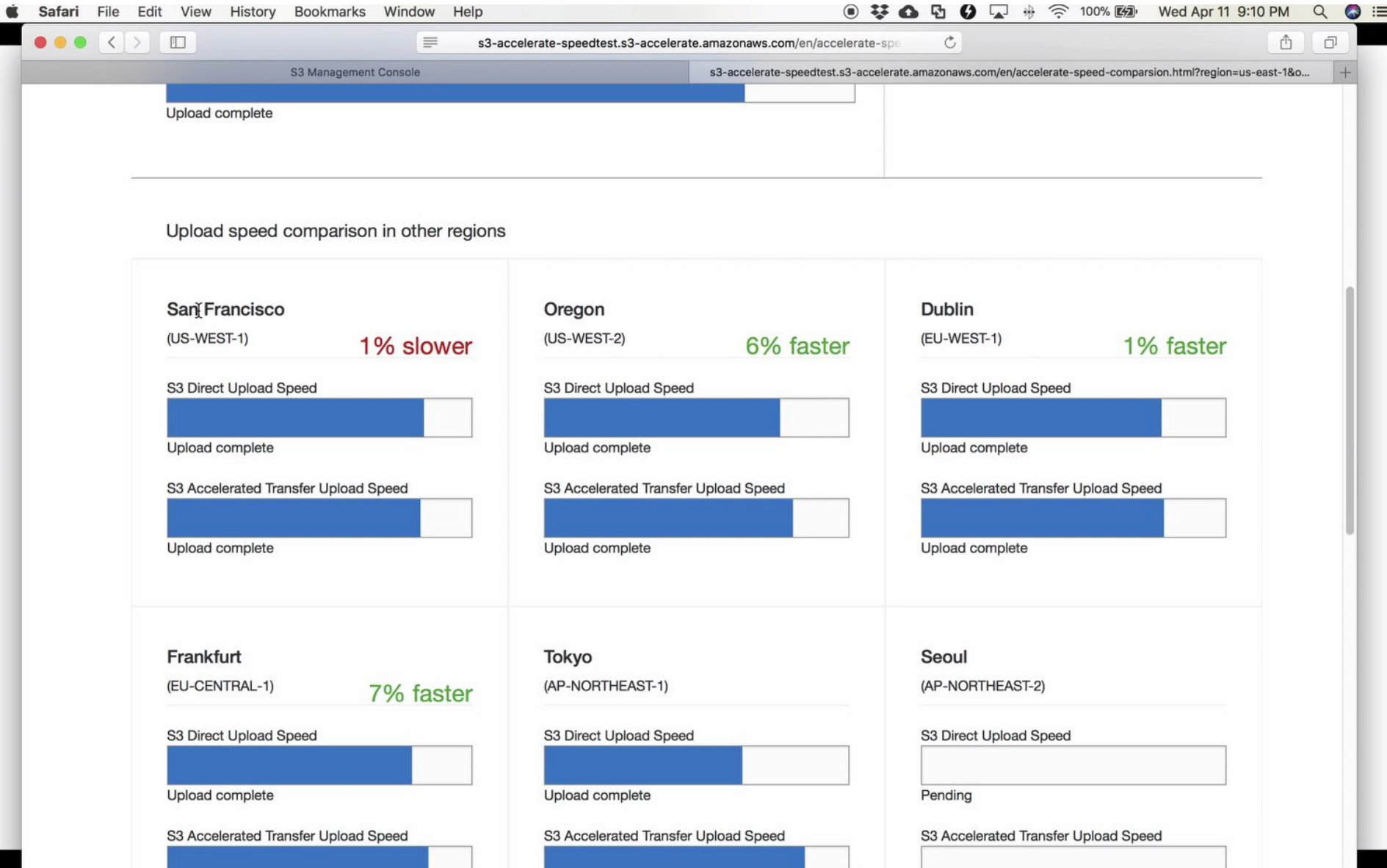
**Events**  
Receive notifications when specific events occur in your bucket.  
[Learn more](#)  
0 Active notifications

**Requester pays**  
The requester (instead of the bucket owner) will pay for requests and data transfer.  
[Learn more](#)  
Disabled

**Transfer acceleration**  
Endpoint: `sanjaykattimani.s3-accelerate.amazonaws.com`  
Use the new accelerated endpoint for faster data transfers, which will incur an additional fee.  
[Want to compare your data transfer speed by region?](#)  
 Enabled  
 Suspended  
Cancel Save



# AMAZON S3 object storage



# AMAZON S3 object storage

The screenshot shows the Amazon S3 console interface. At the top, the browser address bar displays the URL: `s3.console.aws.amazon.com/s3/buckets/sanjaykattimani/?region=us-east-1&tab=pr`. The main content area is titled "Advanced settings" and contains several configuration cards:

- Tags:** A modal window for adding tags. It has a blue header with "Tags" and a close button. It contains two input fields labeled "Key" and "Value", a "+ Add tag" button, and "Cancel" and "Save" buttons at the bottom.
- Transfer acceleration:** A card with a lightning bolt icon. It says "Enable fast, easy and secure transfers of files to and from your bucket." with a "Learn more" link. Below, it shows a checked checkbox for "Endpoint:" with the value `sanjaykattimani.s3-accelerate.amaz...`.
- Events:** A card with a calendar icon. It says "Receive notifications when specific events occur in your bucket." with a "Learn more" link. At the bottom, it shows a radio button selected for "0 Active notifications".
- Requester pays:** A card with a person icon. It says "The requester (instead of the bucket owner) will pay for requests and data transfer." with a "Learn more" link. At the bottom, it shows a radio button selected for "Disabled".

# Cloud Storage security and Integrity

- How to create secure and reliable data storage and access facilities
- Cloud storage security is not much different from existing security practices (e.g. TLS, X509 certificates, encryption etc.)
- At least two concerns when using the cloud
  - The users do not want to reveal their data to the cloud service provider (The data could be sensitive information like medical records)
  - The users are unsure about the integrity of the data they receive from the cloud.
  - Within the cloud, more than conventional security mechanisms will be required for data security

From a user perspective Cloud storage is complex distributed and undefined

No One Understand the Cloud!!!!



# Cloud Storage Security

The single largest security concern that most organizations should have.

As with any WAN traffic, any data can be intercepted and modified.

Data can be located anywhere in the cloud provider data centers

Data can be accessed by provider personnel.

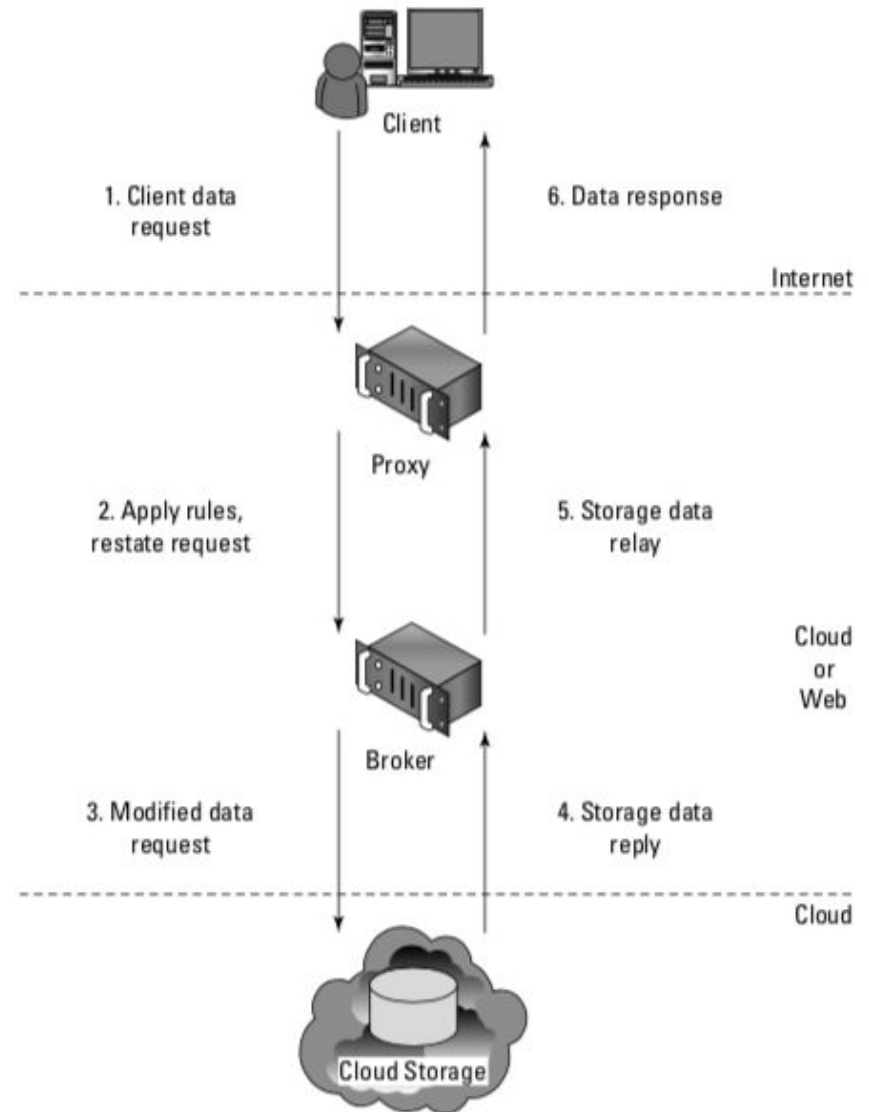
# How to protect my data

- Identify the security boundary separating the client's and vendor's responsibilities
- Determine the sensitivity of the data to risk
- Data should be transferred and stored in an encrypted format.
- Separate clients from direct access to shared cloud storage.
- These are the key mechanisms for protecting data mechanisms:
  - Access control
  - Auditing
  - Authentication
  - Authorizations

# Data Segregation and Isolation

Isolate data from direct client access  
creating a layered access to the data.

Data segregation based on tenants



# Encryption

Most cloud service providers store data in an encrypted form (e.g. Amazon S3 256-bit Advanced Encryption Standard) on server side or client side.

Some example of java code here:

<https://docs.aws.amazon.com/AmazonS3/latest/dev/UsingClientSideEncryption.html>

Problems:

- a problem with encrypted data may result with data that may not be recoverable.
- it does nothing to prevent data loss: keep you keys!!!!



# Encryption example: swift

Encryption of object data at rest on storage node.

What is encrypted:

- Object content i.e. the content of an object PUT request's body
- The entity tag (ETag) of objects that have non-zero content
- All custom user object metadata values i.e. metadata sent using X-Object-Meta- prefixed headers with PUT or POST requests

How:



- Encryption filter in swift proxy server.
  - encrypt object data and metadata when handling PUT and POST requests
  - decrypt object data and metadata when handling GET and HEAD requests

# Encryption example: swift

For security reasons encryption keys are stored using a **Key Management System** (e.g. Barbican, KMIP).

Plaintext data is encrypted to ciphertext using the AES cipher with 256-bit keys implemented by the **python cryptography package**.

only user data is encrypted

Key	Pre-encrypt Values	Key	Post-encrypt Values
Etag	4b7550f00f2e80408b8bb2d6dc7f705f	Etag	LQlpWr6BPR1RUDxmnWrQX1JemA3J egzPI9yd9QmkBOo= 
Content-type	text/plain	Content-type	text/plain
Content-length	28	Content-length	28
X-Object-Meta-Tag	Bank account password	X-Object-Meta-Tag	VEVYRwZYXVVC9QTEFJTg== 
Body	correct horse battery staple	Body	*?/uew(liet#\4*!@j[>.6-fly\$\img alt="lock icon" data-bbox="875 855 915 925"/>

```
X-Object-Meta-Private1: value1
X-Object-Meta-Private2: value2
```

```
X-Object-Transient-Sysmeta-Crypto-Meta-Private1:
  E(value1, object_key, header_iv_1); swift_meta={"iv": header_iv_1,
                                                  "cipher": "AES_CTR_256"}
X-Object-Transient-Sysmeta-Crypto-Meta-Private2:
  E(value2, object_key, header_iv_2); swift_meta={"iv": header_iv_2,
                                                  "cipher": "AES_CTR_256"}
```

```
X-Object-Sysmeta-Crypto-Body-Meta:
  {"iv": body_iv,
   "cipher": "AES_CTR_256",
   "body_key": {"key": wrapped_body_key,
                "iv": body_key_iv}}
```

# Swift: encrypting data on client side using python

## Simple python script to encrypt data

```
from cryptography.fernet import Fernet
key = Fernet.generate_key()
f = Fernet(key)
with open(sys.argv[1], 'r') as input_file, open(sys.argv[2],
'w') as output_file:
    for line in input_file:
        output_file.write(f.encrypt(line) + '\n')
print 'Key: ' + key
```

## upload data on swift as object:

```
swift post -H "X-Delete-After: 5" -m "Some: value"
my_data.crypt
```

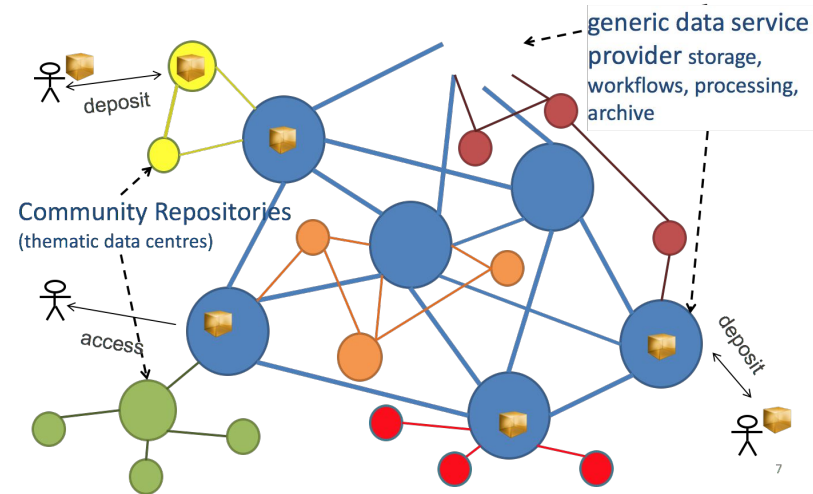
# From Cloud Storage to a Data Cloud

Cloud Storage is an online repository of data.

To build a Data Cloud you need to make data

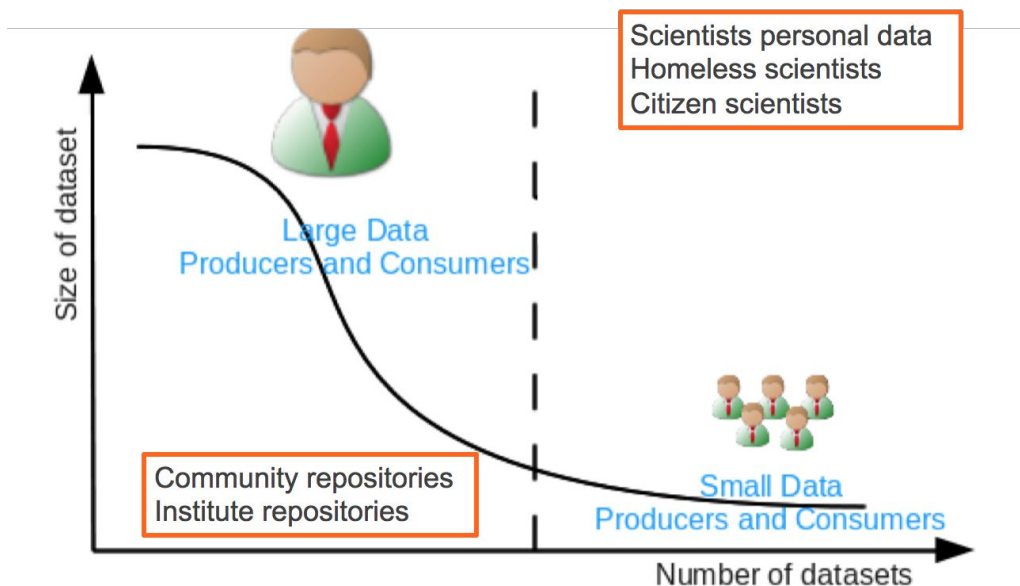
- easy to find,
- easy to share
- easy to process

Data cloud is a distributed infrastructure



# Eudat: a data cloud example

EUDAT is an european Distributed Data Infrastructure (now EOSC) that is used by both scientific projects and single scientists



# EUDAT infrastructure design



**B2DROP**  
Sync and Exchange Research Data



**B2SHARE**  
Store and Share Research Data



**B2SAFE**  
Replicate Research Data Safely



**B2STAGE**  
Get Data to Computation



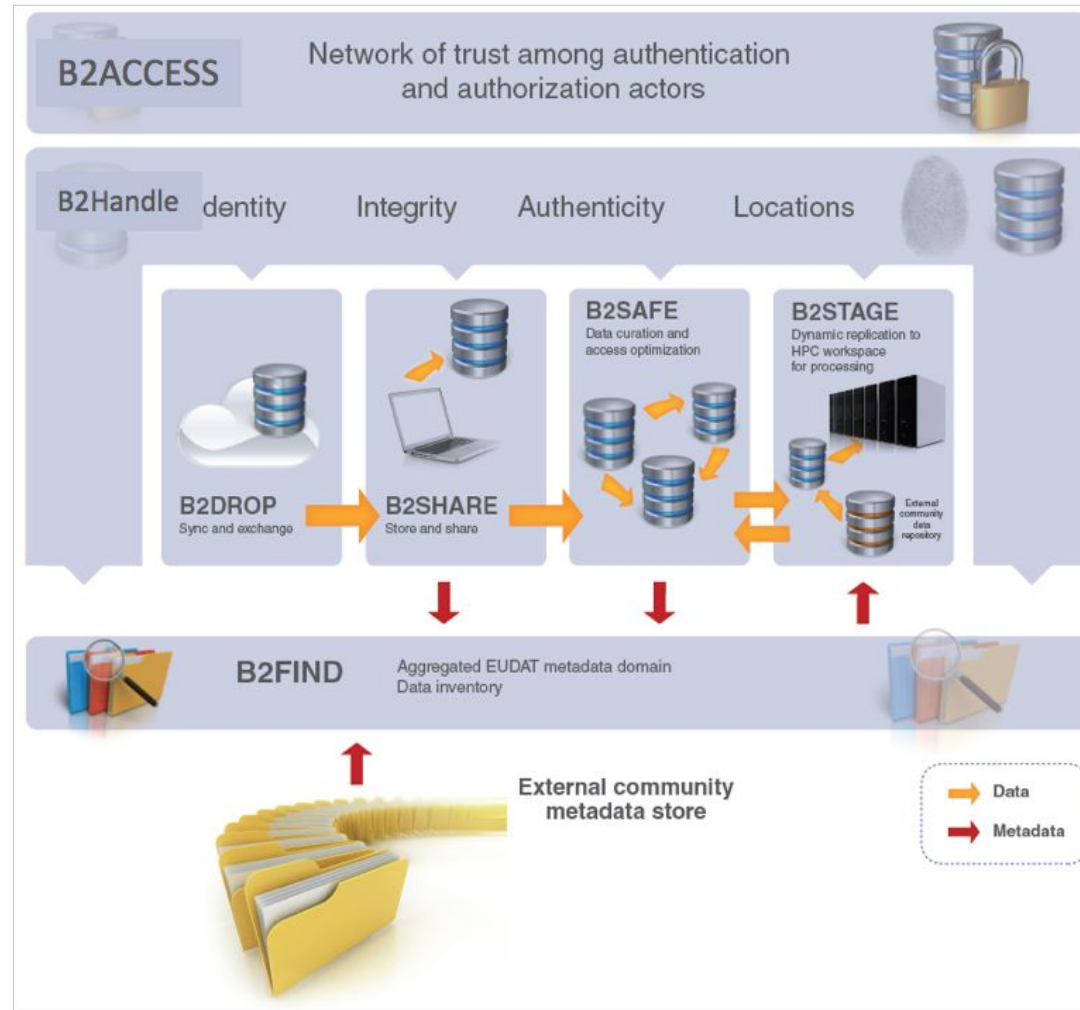
**B2FIND**  
Find Research Data



**B2HANDLE**  
Register your Research Data



**B2ACCESS**  
Identity & Authorisation



# B2DROP

## Who

Citizens Scientists and small teams

## What

**Store and exchange data**

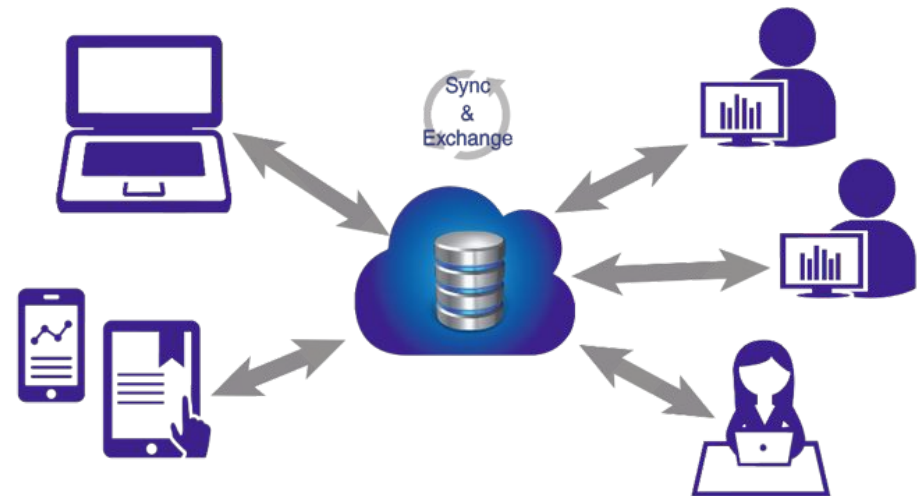
**Synchronize** multiple versions

Ensure **automatic desktop** synchronization

## Why

Ease of Use

Trusted European Service



Integration with **B2ACCESS** to enable access by many different Identity Providers

**Cloud Storage Federation**, collaboration with GEANT in OpenCloudMesh

Assess B2DROP as **workspace** area to **computing facilities**

Name	Size	Modified
EUDAT-ESA-05.02.2015.pptx	6.9 MB	last month
EUDAT - Open D... .pdf	3.8 MB	5 months ago
EUDAT - Open Data-reetz-01.pptx	10.4 MB	5 months ago
EUDAT - Open Data-reetz-01.pptx	14.5 MB	last month





WHAT IS B2DROP    FAQs    CONTACT

Files ▾



Johannes Reetz

All files



New



Name ▲

Size

Modified



**documents**

450 MB

13 minutes ago



**EUDAT**

48.6 MB

last month



**EUDAT2020**

Mark van de Sanden

30.5 MB

5 hours ago



**EUDAT T2.4.3 Workspace**

Magchiel magchiel.bijsterbosch@surfsara.nl

6.6 MB

18 days ago



**reactions\_call-for-interest**

Dieter Van Uytvanck

757 kB

2 months ago



**registry**

Shared

10.5 MB

7 months ago

Deleted files

6 folders

546.9 MB



Course title

Lecture short title

# B2SHARE

## Who

Small to Medium Teams

## What

**Store** data (incl. software) and add domain meta data

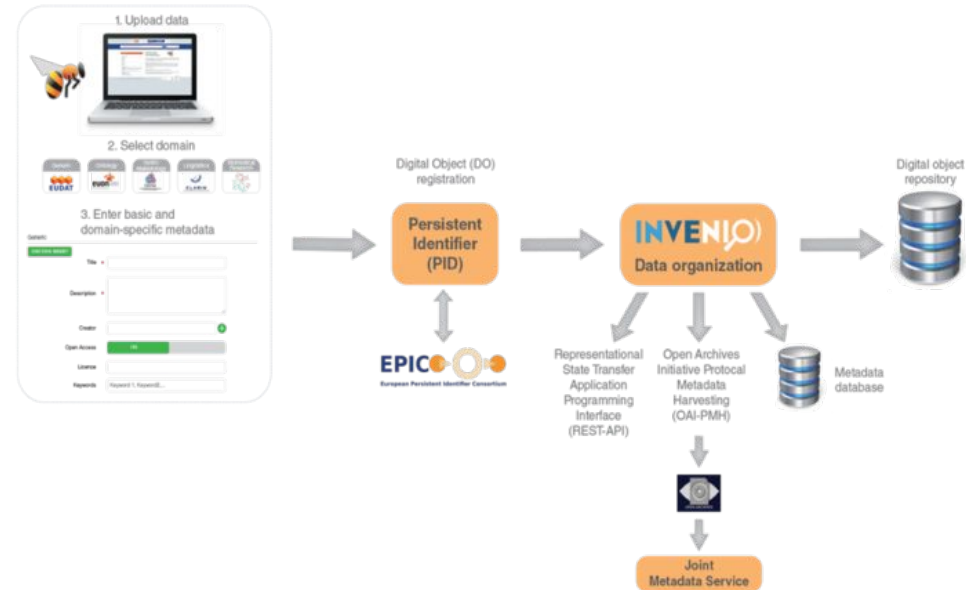
**Share** registered research data worldwide

**Preserve** (small-scale) research data for long-term

## Why

Register Data for Publications

Make known to wider community



Select a domain or project

Linguistics  
Herbarium  
Biodiversity  
Generic  
Ontology  
RDA  
Hydro-Meteorology  
Biomedical Research

Generic

Title \* Title of the resource

Description \*

Creator author

Open Access

Embargo Till

Licence

Keywords keyword1, keyword2, ...

Contact Email contact email

Discipline None selected

Further integration with EUDAT CDI (e.g. B2DROP, B2SAFE)

Integration with **B2ACCESS** (incl eduGAIN), focus on

**authorization**

**Embargo** period

**Editing of metadata**

Data **versioning** and **annotation**

Extended HTTP Restful API interface

Easy installable **software package**



Search 12 records for [ + - ] [ Settings ] [ Search ]



### RDA Governance Document

The role of the RDA Governance Document is to describe the structures of RDA, and their relationships, to support the activity and principles of RDA, and Council's powers and authority. The Governance Document is the responsibility of the RDA Council.

by [Research Data Alliance Council](#) | 18 Sep 2015, 14:52 | [Similar records](#) | [RDA](#) [Policy](#) [Governance](#)

### Data Type Registries working group output; 1

the outcomes of the DTR WG can be summarized as: • Confirmation that detailed and precise data typing consideration in data sharing and reuse and that a federated registry system for such types is highly desirable • needs to accommodate each community's own requirements • Deployment of a prototype registry implementation • one potential data model, against which various use cases can be tested • Involvement of multiple ongoing scientific data management efforts, across a variety of domains, in actively planning for and testing the use of types and associated registries in their data management efforts • Integration with one additional RDA Working Group (Persistent Identifier Types) and at least one Interest Group (RDA/CODATA Materials Data, Infrastructure and Interoperability IG) • Development of a set of questions that require further consideration before a detailed recommendation on data typing can be issued.

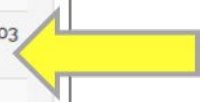
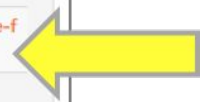
by [Larry Lannom](#) ; [Daan Broeder](#) ; [Girdhar Manepalli](#) ; [Laura Bartolo](#) ; *et al* | 14 Aug 2015, 10:19 | [Similar records](#) | [RDA Recommendation](#) [Data Typing](#) [Registries](#)

**Export**

Export as [BibTeX](#), [MARC](#), [MARCXML](#), [DC](#), [EndNote](#), [NLN](#), [RefWorks](#)

**Metadata**

<b>PID:</b>	<a href="http://hdl.handle.net/11304/42c2e64e-f95e-11e4-8a18-f31aa6f4d448">http://hdl.handle.net/11304/42c2e64e-f95e-11e4-8a18-f31aa6f4d448</a>
<b>Publication:</b>	<a href="http://b2share.eudat.eu">http://b2share.eudat.eu</a>
<b>Publication Date:</b>	2015-05-13
<b>Licence:</b>	Creative Commons Attribution (CC-BY)
<b>Resource Type:</b>	Text
<b>Uploaded by:</b>	enquiries@rd-alliance.org
<b>Contact email:</b>	enquiries@rd-alliance.org
<b>Alternate Identifier:</b>	<a href="http://dx.doi.org/10.15497/4B31A348-03BC-45A4-9BA8-ABFB8EF386C9">http://dx.doi.org/10.15497/4B31A348-03BC-45A4-9BA8-ABFB8EF386C9</a>
<b>Domain:</b>	RDA
<b>Coverage:</b>	Official Document
<b>Checksum:</b>	d1a8a9f147c60904a4201e5a8333515ebb2bbf97c91e54cc0a0df9825bebf9c7





**B2SHARE**  
Store and Share Research Data



**B2DROP**  
Sync and Exchange Research Data



GO TO EUDAT WEBSITE

**Select B2DROP files**

File name	Size	Date
EUDAT2020		10/15/2015, 3:11:32 PM
EUDAT2020 KO 2015		4/2/2015, 4:01:08 PM
documents		9/5/2014, 10:39:14 PM
Conference Party		10/1/2014, 12:19:45 AM
EUDAT2 Graphics		2/22/2015, 11:38:27 PM
IRODS workshop		4/15/2015, 10:56:56 AM
<input type="checkbox"/> _DS_Store	6 KB	9/2/2015, 3:53:02 PM
<input type="checkbox"/> .._DS_Store	4 KB	9/2/2015, 3:52:07 PM
<input type="checkbox"/> Bz Services integration.jpg	68 KB	9/2/2015, 3:52:52 PM
<input type="checkbox"/> ownCloudUserManual.pdf	2 MB	9/5/2014, 9:37:06 PM
<input type="checkbox"/> EUDAT-B2DROP.pdf	61 MB	9/17/2014, 1:08:21 PM

Cancel Select

Powered by **INVENIO**

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065. Legal Notice

Terms of Use REST API About EUDAT

GO TO EUDAT WEBSITE

Step 01

Drag and drop files here

Select files

Start upload

Filename	Size	Status	Remove
ELIXIR EBI Subscription Data Distribution use case v3.png	101 kb	<div style="width: 100%;"></div>	

Step 02

Select a domain or project

Linguistics

Herbarium

Biodiversity

Generic

Ontology

RDA

Hydro-Meteorology

Biomedical Research

Deleted files

# B2SAFE

## Who

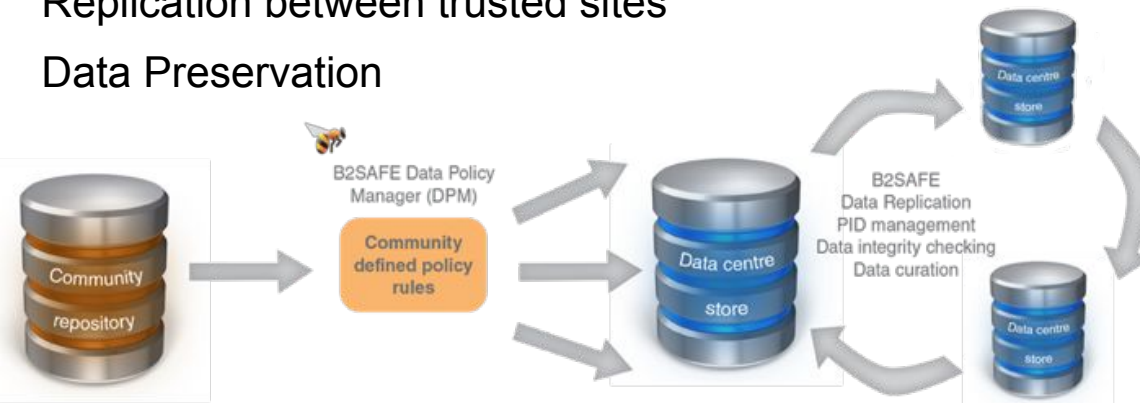
Community Data Managers  
'Sophisticated' Organisations

## What

Provide an abstraction layer which virtualizes large-scale data resources  
Guard against data loss in long-term **archiving and preservation**  
**Optimize access** for users from different regions  
Bring data **closer to powerful computers**

## Why

Performance  
Replication between trusted sites  
Data Preservation



Support **iRODS v4**  
Support **metadata**  
Optimize and **extend** policies to support **data curation** and **provenance**  
Further integration with **B2ACCESS**  
Support **authorization** on basis of **community access rules**  
Assess B2SAFE as **workspace** area to **computing facilities**



# B2SAFE

Data policies are **centrally managed**

Policy rules are implemented and enforced by **site-local rule engines**

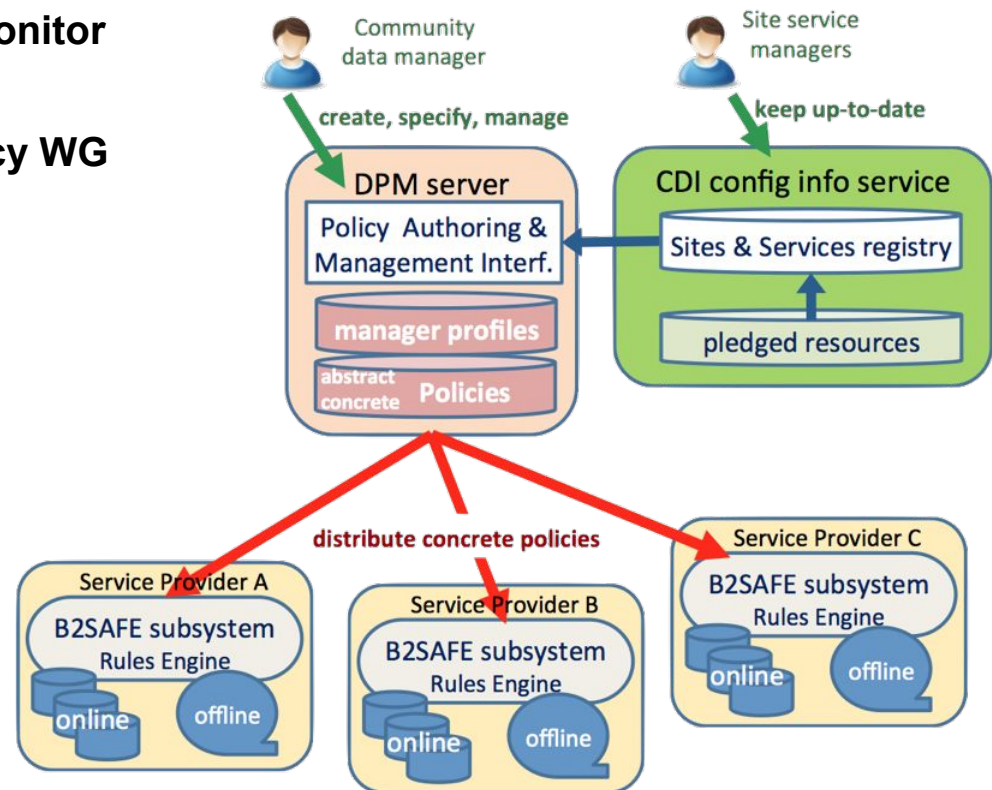
Policies describe in an **abstract** language

Community data managers must **authenticate** to provide **trust**

Support policies for **data replication** and **integrity checking**

**Central logging** for **auditable** data policies to **monitor** execution

Active collaboration with the **RDA Practical Policy WG**



# B2STAGE

## Who

Users and Communities with Significant Computational Needs

## What

**Transfer** large data collections from EUDAT storages to **external HPC facilities for processing**

**Copy large data sets, ingesting them** onto EUDAT storage resources

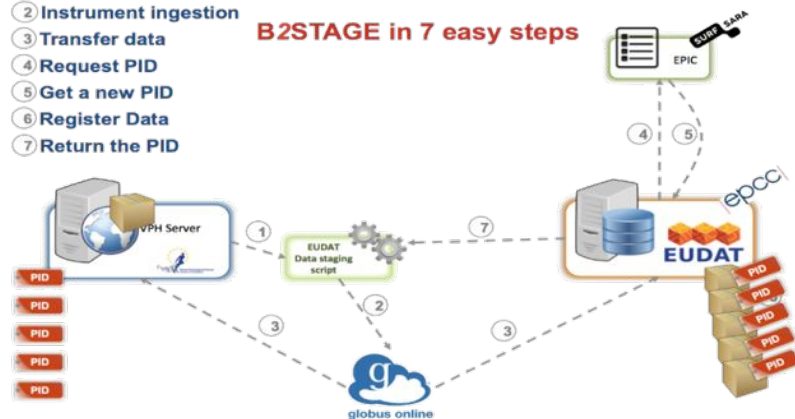
## Why

Integration/Collaboration with PRACE

Simplify Data Transfer

- 1 Ingest new data
- 2 Instrument ingestion
- 3 Transfer data
- 4 Request PID
- 5 Get a new PID
- 6 Register Data
- 7 Return the PID

**B2STAGE in 7 easy steps**



Further develop **HTTP** to a **mature interface** and extend functionality to **metadata**

**Native support PIDs** within **GridFTP** transfers

Extend EUDAT **client API** library to other **B2 services** (e.g. B2SHARE, B2FIND, PID)  
Further integration with B2ACCESS

# B2FIND

## Who

Anyone

## What

**Find** collections of scientific data quickly and easily, irrespective of their origin, discipline or community

Get quick **overviews** of available data

Browse through collections using **standardized facets**

## Why

Unique collection

Ease of Searching

**Community customizations**

**Annotation** of datasets

Further assess **RDF** and **Linked Data**



Search communities...

13 communities found Order by: Name Ascending

 <p><b>ALEPH</b> ALEPH was a particle physics experiment installed on the Large... 185 Datasets</p>	 <p><b>B2SHARE</b> EUDAT offers B2SHARE - a user-friendly, secure, robust, reliable and... 182 Datasets</p>	 <p><b>CESSDA</b> CESSDA is an umbrella organisation for the European national data archives.... 529 Datasets</p>
 <p><b>CLARIN</b> The Common Language Resources and Technology Infrastructure CLARIN project is... 137451 Datasets</p>	 <p><b>DataCite</b> DataCite is a not-for-profit organisation formed in London on 1 December... 241180 Datasets</p>	 <p><b>EARLINET</b> EARLINET, the European Aerosol Research Lidar Network, (www.earlinet.org) is... 5 Datasets</p>
 <p><b>GBIF</b> The Global Biodiversity Information Facility (GBIF) was established by... 13814 Datasets</p>	 <p><b>ENES</b> The European Network for Earth System modelling (ENES) provides information... 830 Datasets</p>	 <p><b>NARCIS</b> NARCIS provides access to scientific information, including (open access)... 175442 Datasets</p>
 <p><b>PDC</b> The PDC is a database of metadata and data that describes, indexes, and... 1959 Datasets</p>	 <p><b>PaNdata</b> PaNdata - the Photon and Neutron data infrastructure initiative - brings... 930 Datasets</p>	 <p><b>SDL</b> Search Digital Libraries (SDL) is hosted by the Documentation Research and... 23903 Datasets</p>
	 <p><b>The European Library</b> The European Library comprises digital resources supplied by a prestigious... 26958 Datasets</p>	



 / Datasets

 Filter by location Clear

 Map data © OpenStreetMap contributors  
 Tiles by MapQuest

 Publication Year Clear
 to 
 Communities

DataCite (2027)

ENES (123)

SDL (7)

CLARIN (3)

PaNdata (1)

NARCIS (1)

B2SHARE (1)

[Show More Communities](#)
 Tags

Dataset (2026)

climate simulation (93)

IPCC-AR4 (87)

scenario run (84)

ECHAM5 (51)

ipcc 
**2,163 datasets found for "ipcc"**
Order by: Relevance 
**IPCC DDC AR5 ETH Zurich/IPCC AR5 WGI esm experiments**

\*esmX\* are an experiment family of the CMIP5 - Coupled Model Intercomparison Project Phase 5 (<http://cmip-pcmdi.llnl.gov/cmip5/>) including esmHistorical, esmrcp85, esmFdbk1,...

**IPCC DDC AR5 ETH Zurich/IPCC AR5 WGI decadal experiments**

\*decadalYYYY\* are an experiment family of the CMIP5 - Coupled Model Intercomparison Project Phase 5 (<http://cmip-pcmdi.llnl.gov/cmip5/>). CMIP5 is meant to provide a framework...

**IPCC DDC AR5 ETH Zurich/IPCC AR5 WGI additional historical experiments**

\*historicalXXX\* are an experiment family of additional historical experiments of the CMIP5 - Coupled Model Intercomparison Project Phase 5 (<http://cmip-pcmdi.llnl.gov/cmip5/>)...

**IPCC DDC AR5 ETH Zurich/IPCC AR5 WGI noVole experiments**

\*noVoleYYYY\* are an experiment family of the CMIP5 - Coupled Model Intercomparison Project Phase 5 (<http://cmip-pcmdi.llnl.gov/cmip5/>). CMIP5 is meant to provide a framework...

**IPCC-DDC\_CSIRO\_SRES\_A1\_TMP**

Abstract | Project: IPCC Data Distribution Centre : Third Assessment Report data sets The Intergovernmental Panel on Climate Change (IPCC) has been established by WMO und...

**IPCC-DDC\_CSIRO\_SRES\_A1\_TMP850**

Abstract | Project: IPCC Data Distribution Centre : Third Assessment Report data sets The Intergovernmental Panel on Climate Change (IPCC) has been established by WMO und...

**IPCC-DDC\_CSIRO\_SRES\_A1\_TMP500**

Abstract | Project: IPCC Data Distribution Centre : Third Assessment Report data sets The Intergovernmental Panel on Climate Change (IPCC) has been established by WMO und...

**Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS-3 ...**

The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS) set is a completely satellite based climatology of precipitation, turbulent heat fluxes and...

 1 2 3 ... 19204 

# B2HANDLE

## Who

Groups or Communities who want to make their data citable

## What

Follows **policies** to register data and make it **long term refer-** and **citable**

Reliability through mutual **PID mirroring**

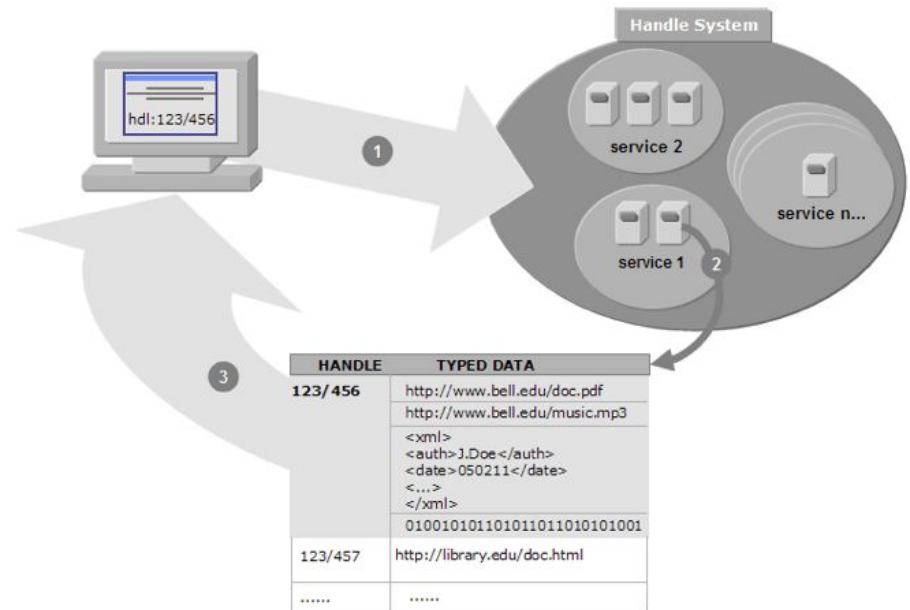
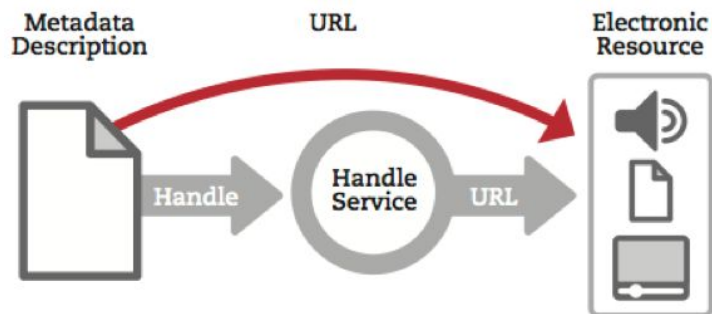
Provides **abstraction layer** between a globally **unique persistent identifier** and **physical location** of data objects

Machine readable via **HTTP RESTful API**

## Why

Simple integration

Technology Agnostic



# B2ACCESS

## Who

Anyone wanting to use the B2 Services

## What

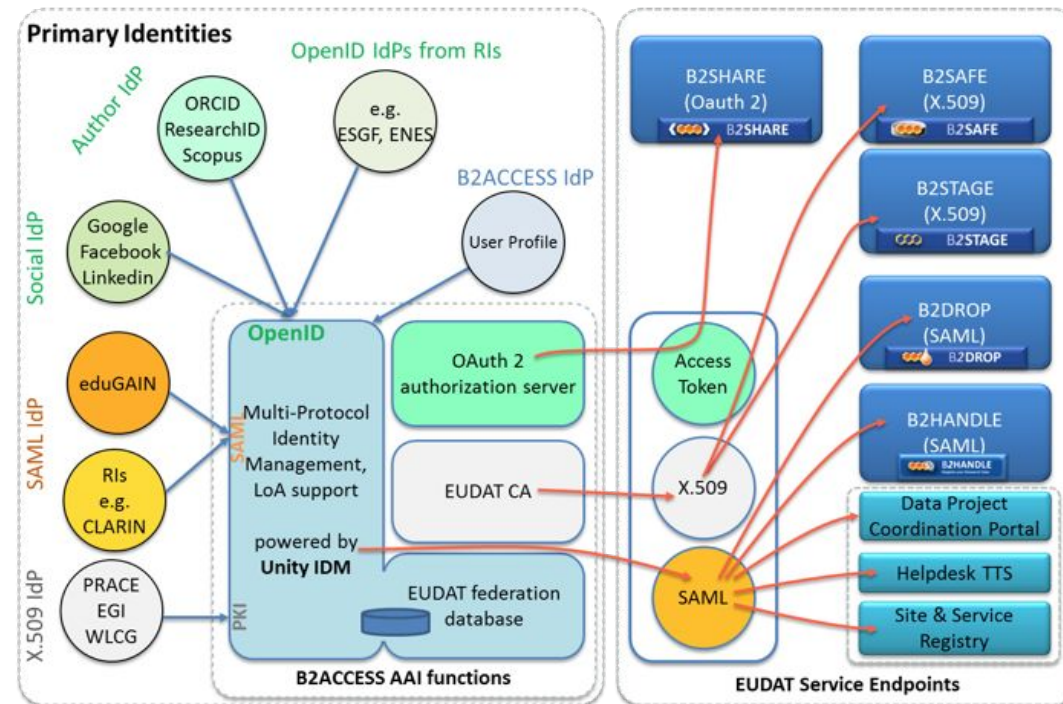
Complies with **community ownerships** and **access rights**, basis of trust

Credential **conversion approach** (e.g. SAML, OpenID, X.509, Username/password)

Identity provider for **citizen scientists**

## Why

Use your own ID in federated environment





# Login to B2ACCESS

[Register a new account](#)

Username:

Password:

Authenticate




## Select your primary identity information provider


Search:

Log in with your B2ACCESS ID







[Password](#) [Certificate authentication](#)

Log in with your social ID



Log in with your organisational ID

-  Academy of Performing Arts in Prague
-  Academy of Sciences Library
-  Academy of Visual Arts Leipzig
-  Academy of fine arts Dresden
-  Accrington & Rossendale College
-  Activate Learning

# Eudat APIs list

<https://github.com/EUDAT-Training>