

Advanced Coding and Cloud Computation

...

Romolo Politi

Lecture List

Elenco Lezioni

- September 16th 2024

Lecture September 16th 2024

Course Overview

Course Overview

Cloud

- Cloud structure
- Data in the Cloud
- Cloud Computing

Data

- Data and Metadata
- Archives
- Relational and not-relational Database

Computing

- Retrieval
- Manipulation
- Visualization

Environment

- Virtualization and Containers
- Microservices
- DevOps

Coding

- Fundamentals of Coding
- Python
- Versioning and Documentation

Tools

- Slides and Examples available on GitHub:
 - <https://github.com/RomoloPoliti-INAF/PhDCourse2024>
- The example will be written in Python 3.12
- Microsoft Visual Studio Code will be used as framework
 - <https://code.visualstudio.com>

Struttura del Corso

- The list of topics shown earlier was organized by categories.
- We will follow an example-driven approach to better understand the philosophy behind it.
- After the introduction to programming, we will develop an example of a complex program (State Machine).
- Lastly, we will develop a WebApp and prepare it for deployment in containers.
- For some topics, we will not go into detail because the purpose of the course is to provide a general overview of the subject.
- Even though they won't be discussed, many details will be available in the slides or through the provided links.

Cloud Definition

What's Cloud

It is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user.

wikipedia

Cloud Types

In Promise



Out Promise



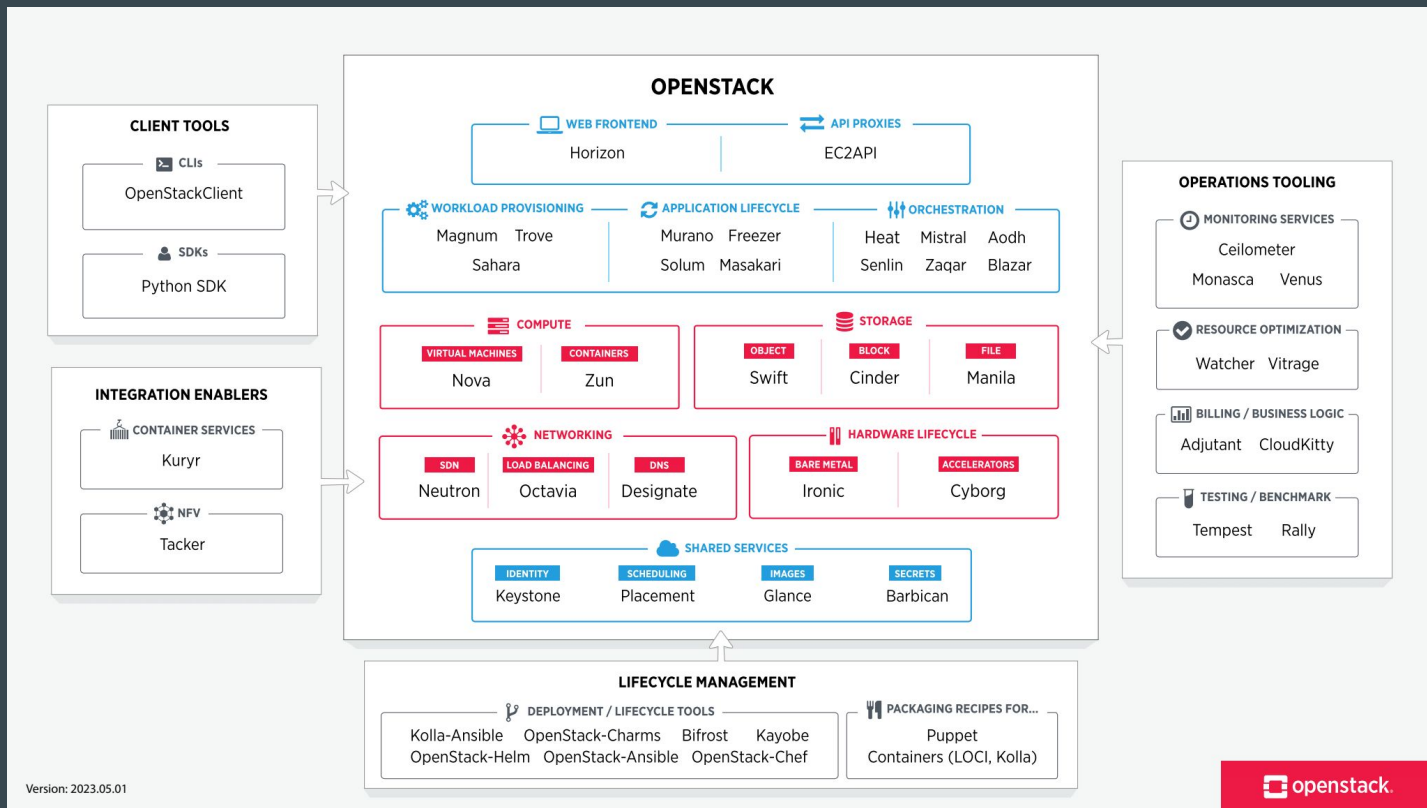
Google Cloud



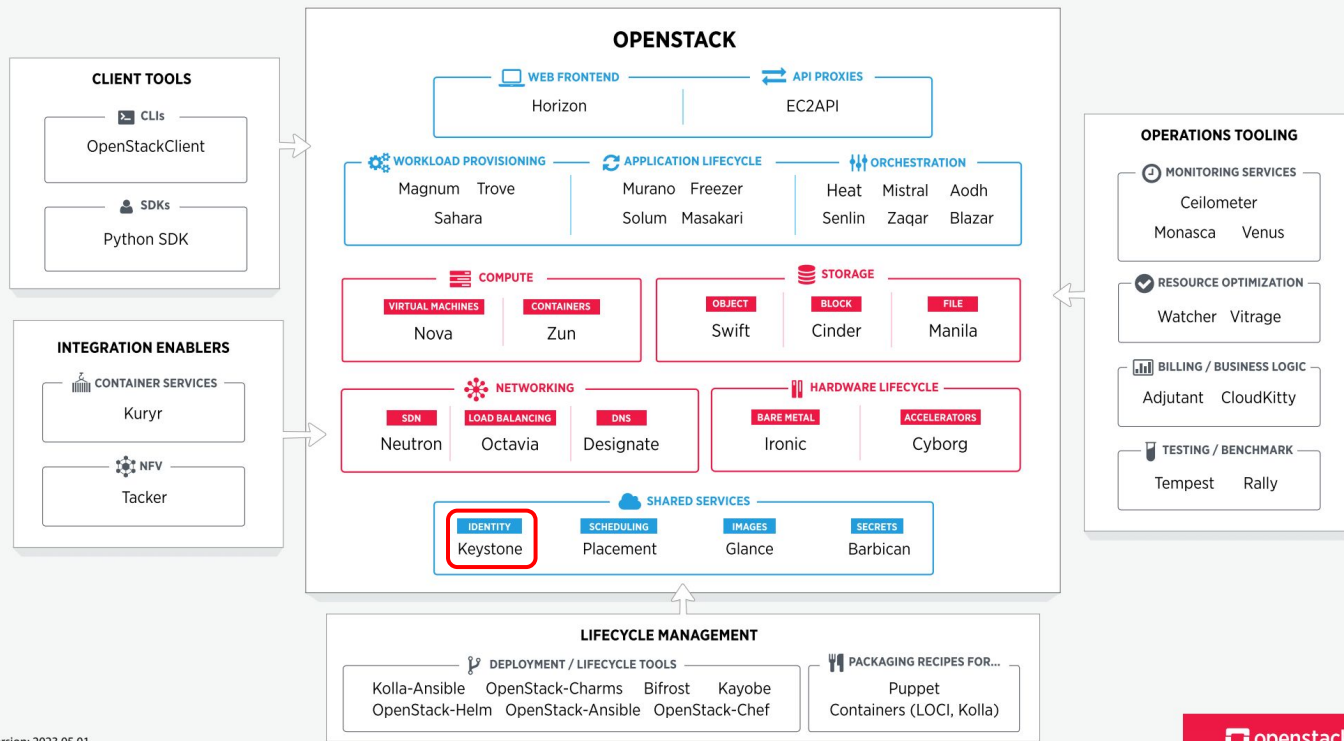
Cloud Structure

Cloud Structure

<https://www.openstack.org/>

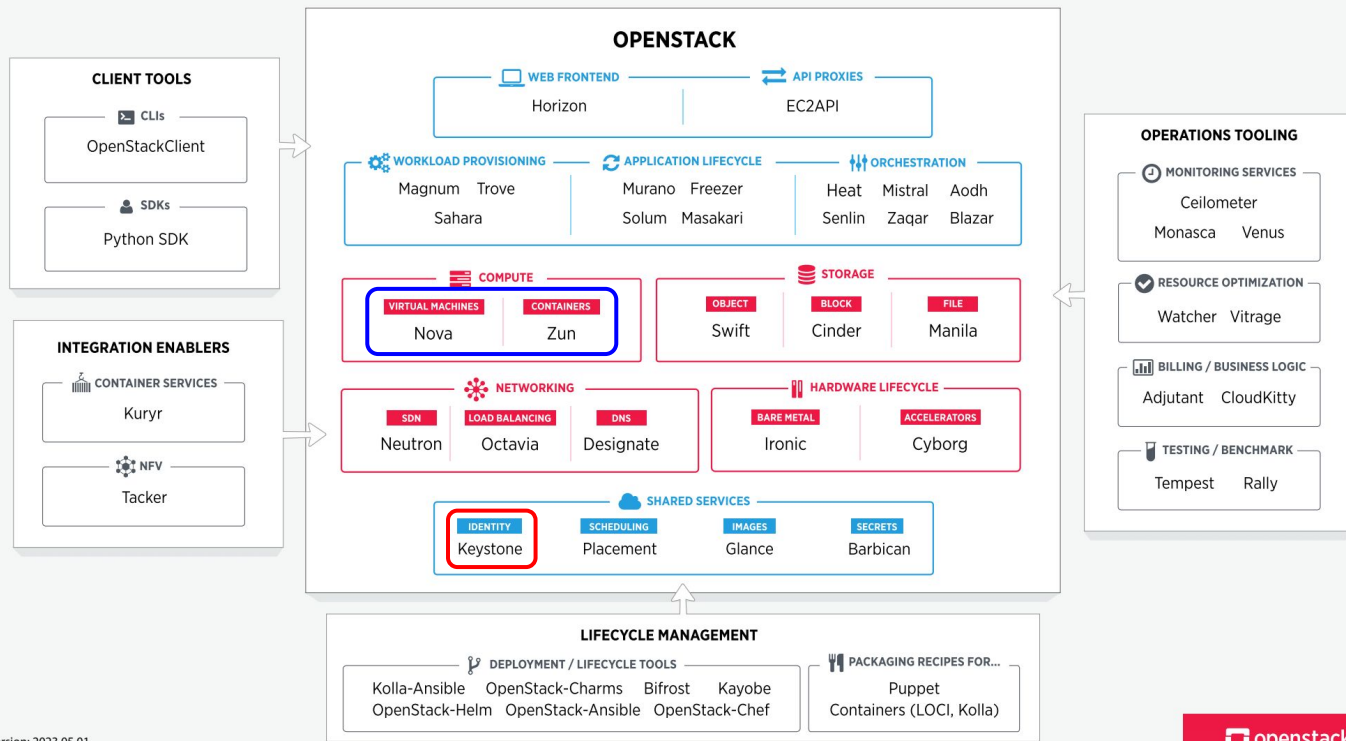


Cloud Structure



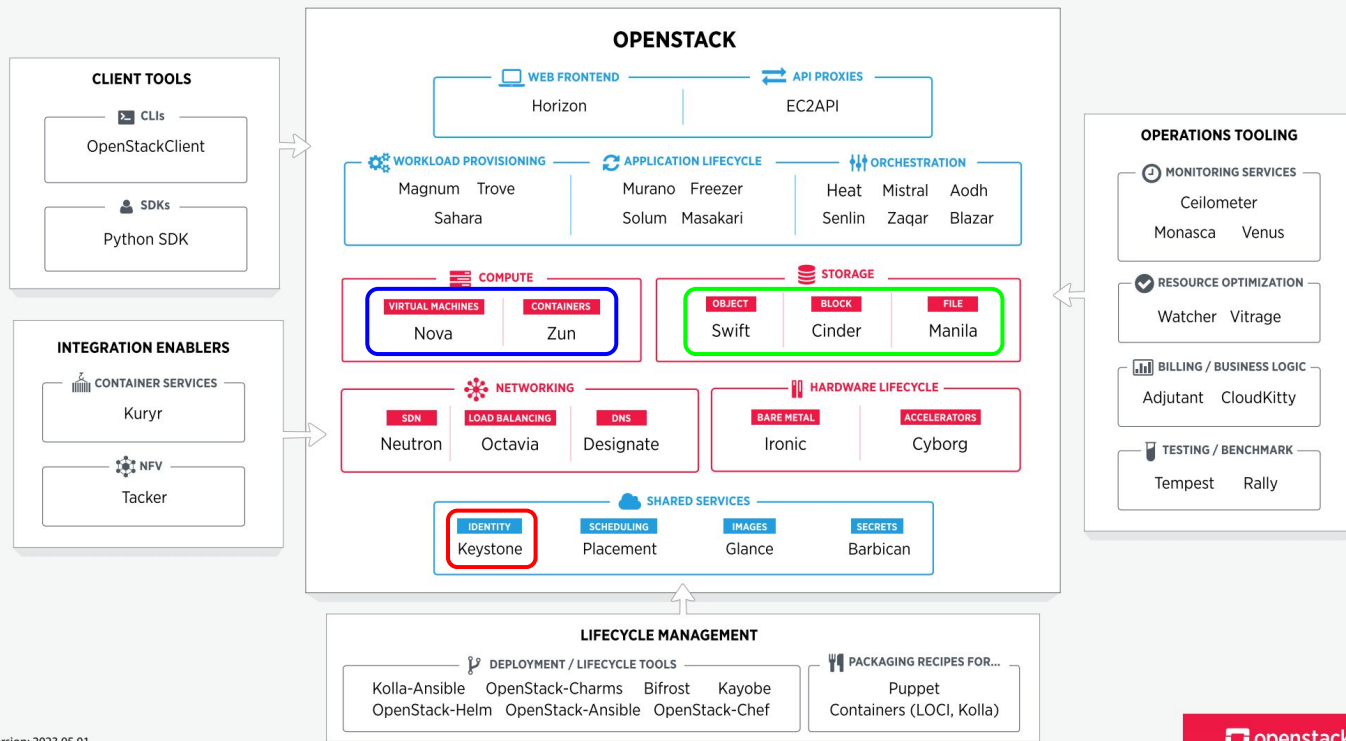
- Identity

Cloud Structure



- Identity
- Compute

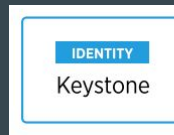
Cloud Structure



- Identity
- Compute
- Storage

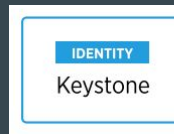
Main Components

- IAM (Identity and Access Management)



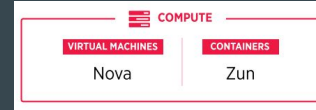
Main Components

- IAM (Identity and Access Management)
 - identity check
 - list of resources
 - privileges
 - credits (cloud off premise)

A screenshot of the Google sign-in interface. At the top is the Google logo. Below it is the text "Sign in with your Google Account". In the center is a large, light gray circular placeholder for a profile picture. Below the placeholder are two input fields: the first is labeled "Email" and the second is labeled "Password". Below these fields is a blue button with the text "Sign In" in white. At the bottom left is a checkbox with the text "Stay signed in", and at the bottom right is a link that says "Need help?".

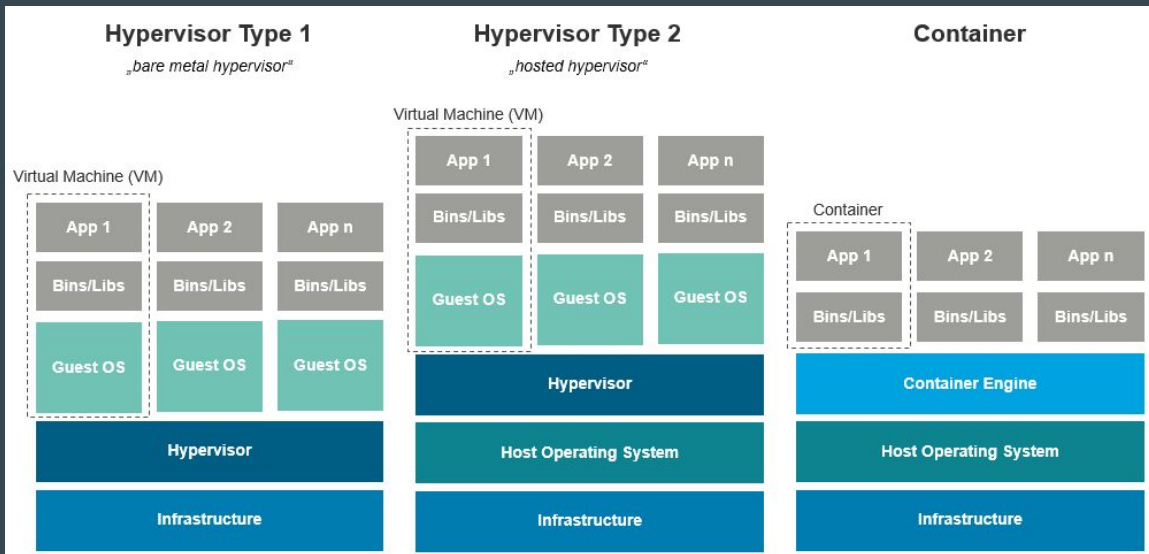
Main Components

- IAM (Identity and Access Management)
 - identity check
 - list of resources
 - privileges
 - credits (cloud off premise)
- Compute Services



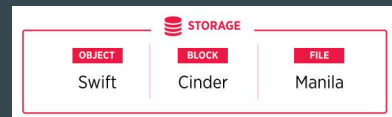
Main Components

- IAM (Identity and Access Management)
 - identity check
 - list of resources
 - privileges
 - credits (cloud off premise)
- Compute Services



Main Components

- IAM (Identity and Access Management)
 - verifica identità
 - lista di risorse dedicate
 - privilegi
 - Credito (cloud off premise)
- Compute Services
- Storage Services

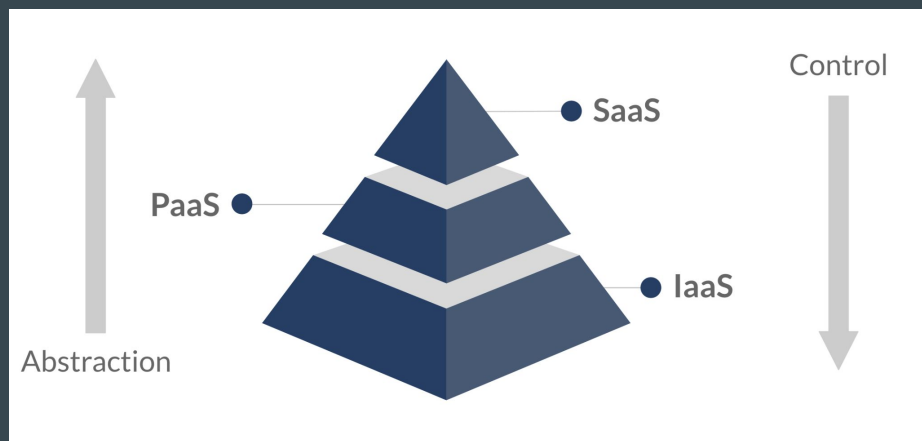


Servizi Cloud

- Infrastructure as a Service
- Platform as a Service
- Software as a Service

Cloud Services

- Infrastructure as a Service
- Platform as a Service
- Software as a Service

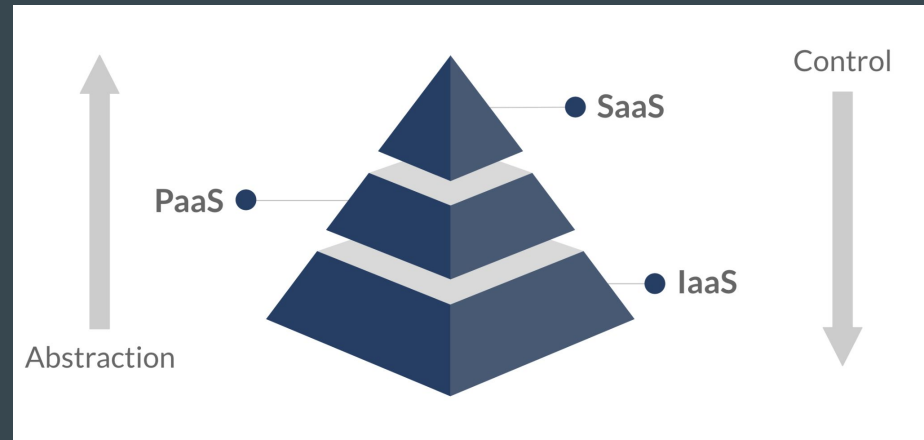


IaaS

provides virtualized computing resources (CPU, RAM, disks, etc.) over the internet, allowing users to manage and scale hardware infrastructure without physical ownership.

Cloud Services

- Infrastructure as a Service
- Platform as a Service
- Software as a Service

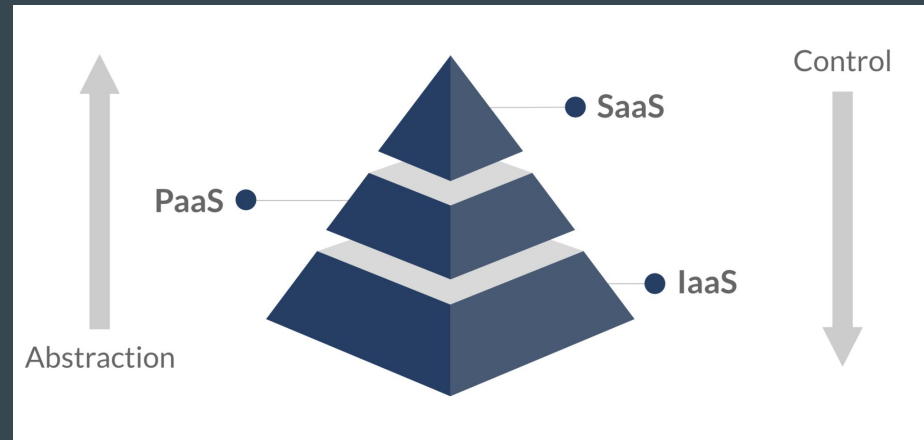


PaaS

It provides a cloud-based environment where developers can build, deploy, and manage applications without dealing with the underlying infrastructure.

Servizi Cloud

- Infrastructure as a Service
- Platform as a Service
- Software as a Service

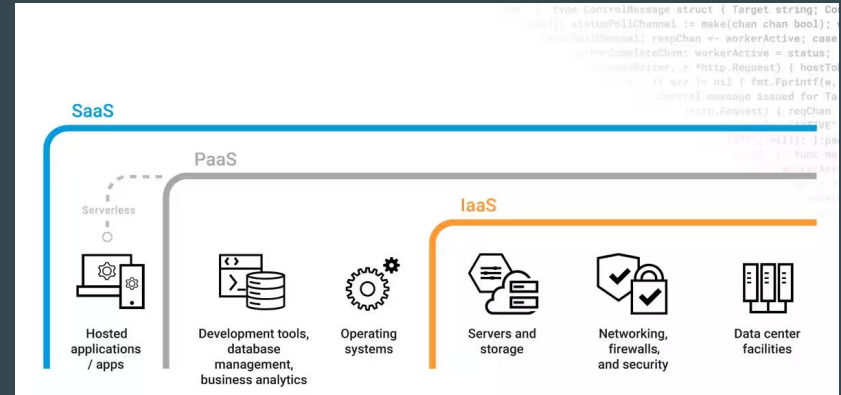
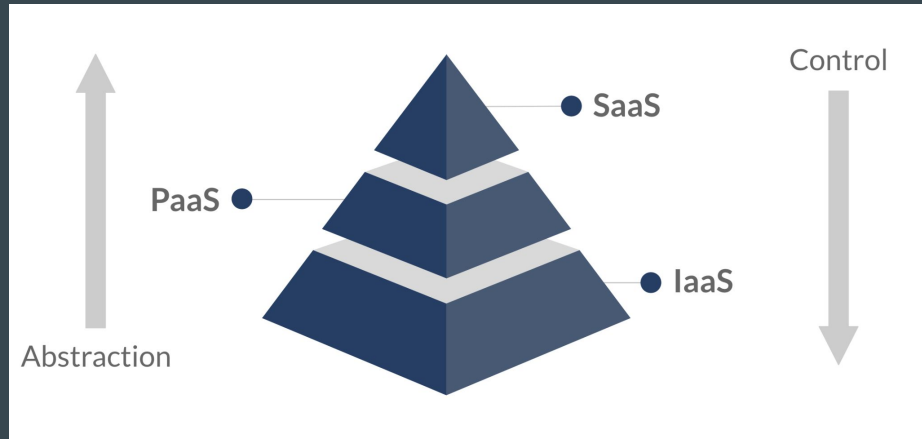


SaaS

it is a cloud-based model where applications are hosted and provided over the internet, allowing users to access and use software without managing the underlying infrastructure.

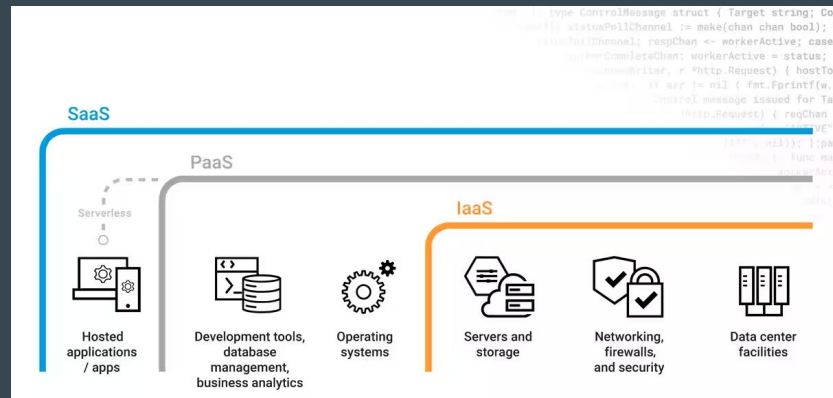
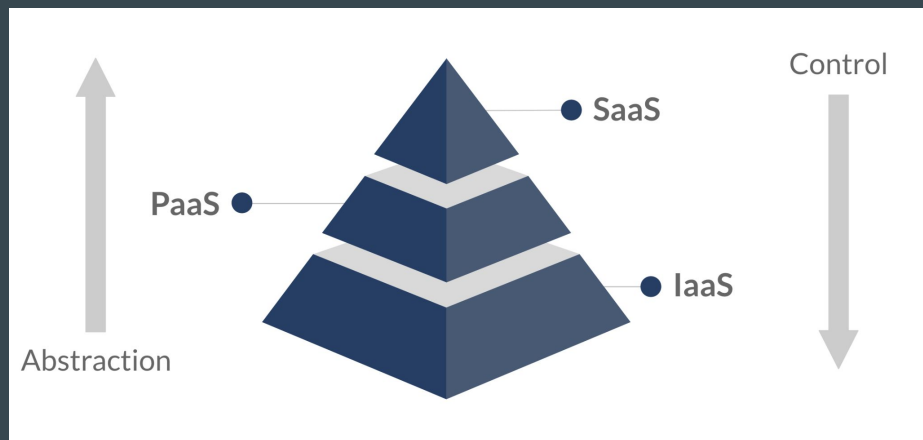
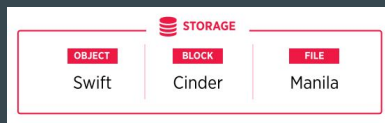
Servizi Cloud

- Infrastructure as a Service
- Platform as a Service
- Software as a Service



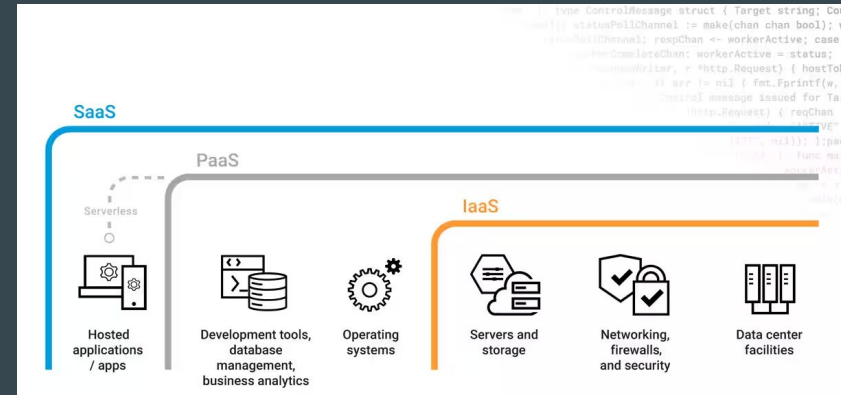
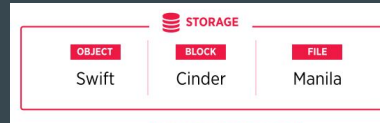
Servizi Cloud

- Infrastructure as a Service
- Platform as a Service
- Software as a Service
- Data as a Service



Servizi Cloud

- Infrastructure as a Service
- Platform as a Service
- Software as a Service
- Data as a Service



Data and Metadata

Data Definition

In computing, a data is a collection of facts, figures, or details that can be processed or analyzed. It often consists of numbers, text, or other types of information that are recorded and stored electronically.

Data serves as the foundation for creating information and insights through analysis and interpretation. It can be raw, unprocessed input or structured and organized to facilitate meaningful conclusions. In various contexts, data is used to make decisions, generate reports, or drive machine learning algorithms. Proper management and understanding of data are crucial for effective decision-making and problem-solving.

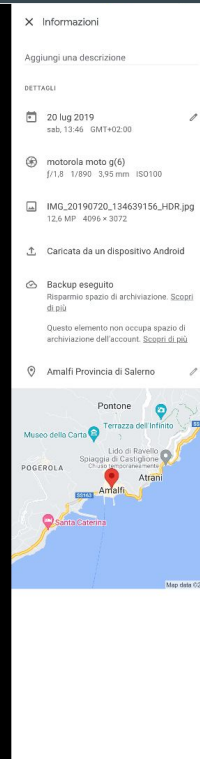
Metadata Definition

A metadata is data that provides information about other data. It describes various attributes of data, such as its origin, format, and relationships to other data, which helps in organizing, managing, and retrieving it efficiently.

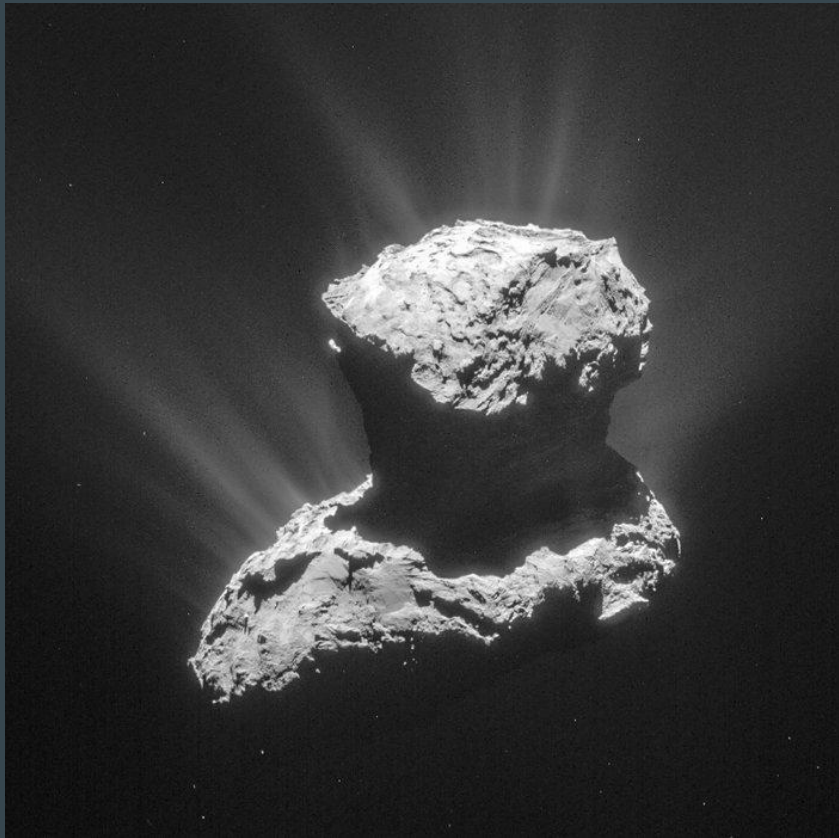
Metadata can include details like the creator of a file, the date it was created, and how it should be used. It is essential for data cataloging and improving the accessibility and usability of information.

By offering context and structure, metadata enhances data searchability and interoperability across different systems and platforms.

A wide-angle photograph of the Duomo di Pistoia, a large Romanesque cathedral. The building features a prominent bell tower on the left and a wide staircase leading to the entrance. The facade is decorated with intricate geometric patterns and a large rose window. The surrounding area includes modern buildings and a clear blue sky.

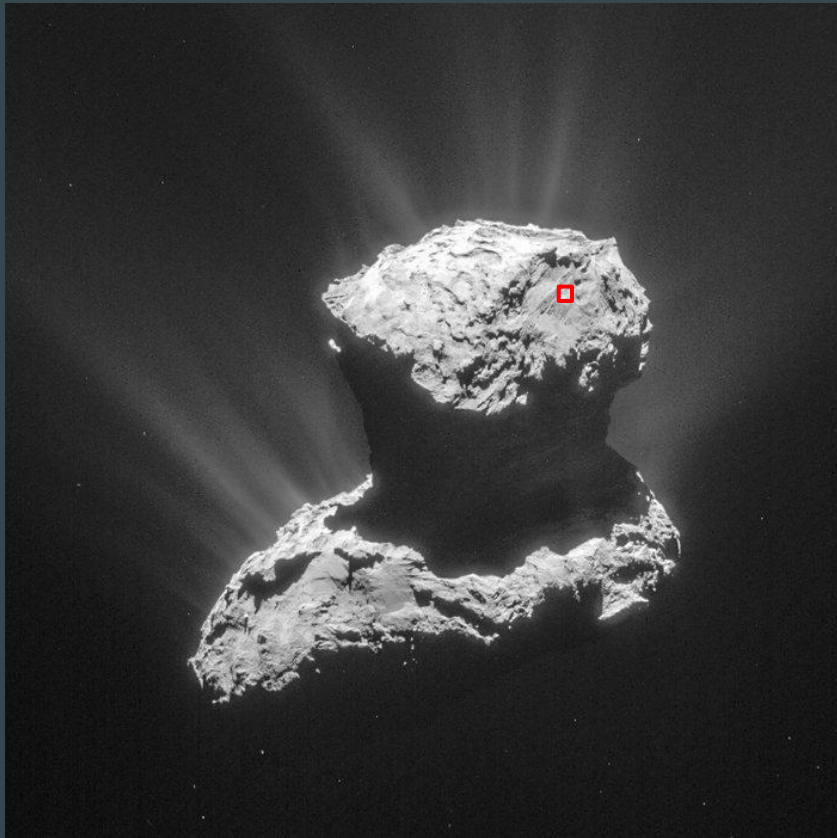


Planetological Example



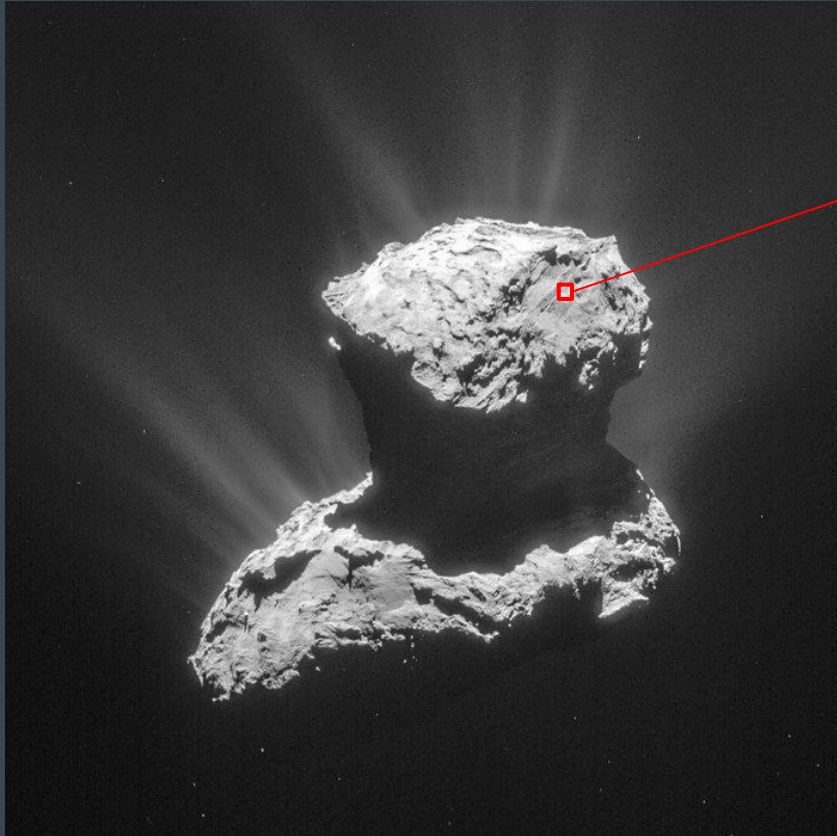
Which is the data?

Planetological Example



Which is the data?

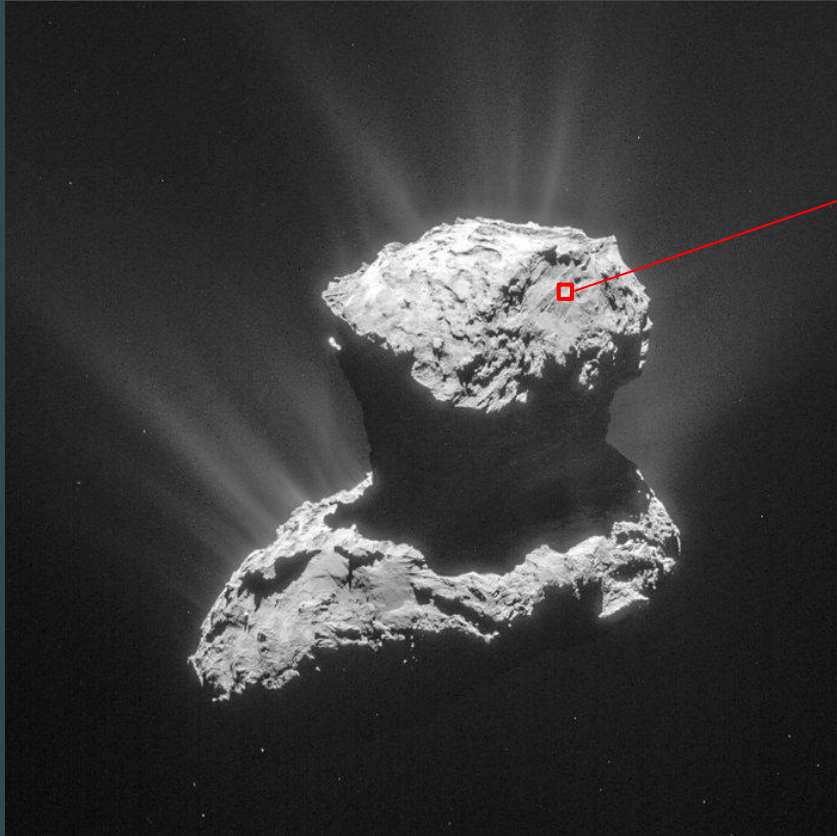
Planetological Example



Which is the data?

The pixel is represented as a 32-bit floating-point number.

Planetological Example

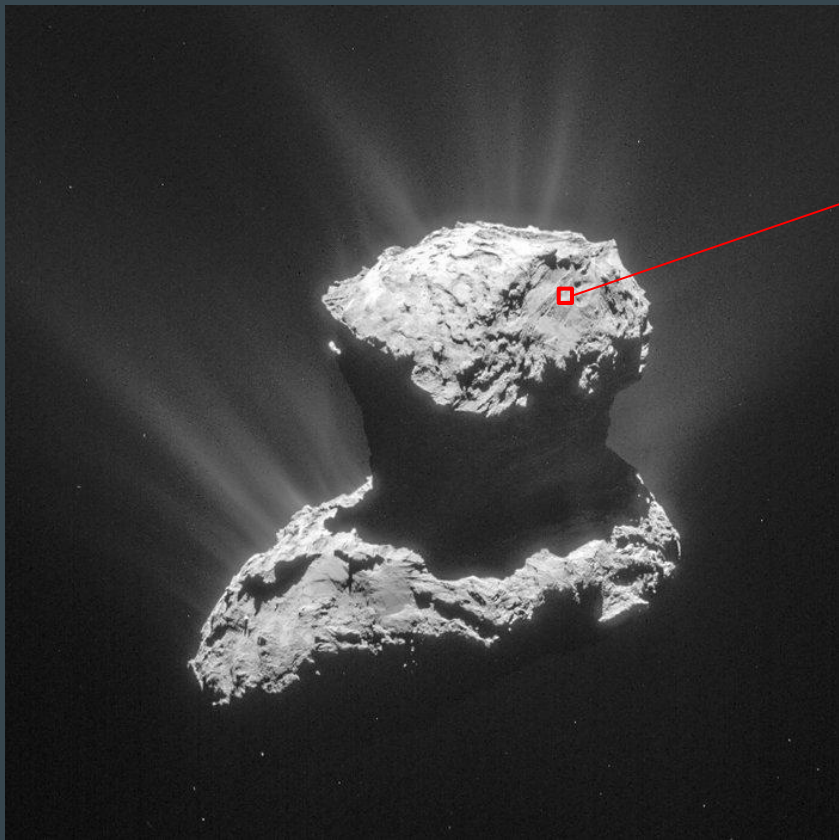


Which is the data?

The pixel is represented as a 32-bit floating-point number.

Does this have meaning?

Planetological Example



Which is the data?

The pixel is represented as a 32-bit floating-point number.

Does this have meaning?

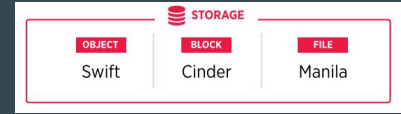
The answer is **NO**.

It is necessary to know

- lighting,
- comet position,
- spacecraft position,
- exposure times,
- acquisition mode,
- pixel georeferencing.

Data in the Cloud

Cloud Storage



Cloud Storage



Block storage divides data into separate components made up of fixed-size data blocks, each with a unique identifier. Block storage allows the underlying storage system to retrieve it regardless of where it is stored.

Cloud Storage



Block storage divides data into separate components made up of fixed-size data blocks, each with a unique identifier. Block storage allows the underlying storage system to retrieve it regardless of where it is stored.

File storage is the most well-known storage format: data is stored in files that can be interacted with, contained in folders within a hierarchical file directory.

Cloud Storage



Block storage divides data into separate components made up of fixed-size data blocks, each with a unique identifier. Block storage allows the underlying storage system to retrieve it regardless of where it is stored.

File storage is the most well-known storage format: data is stored in files that can be interacted with, contained in folders within a hierarchical file directory.

Object storage is a storage format in which data is stored in separate units called objects. Each unit has a unique identifier, or key, that allows it to be located independently of where it is stored in a distributed system.

Cloud Storage



Block storage divides data into separate components made up of fixed-size data blocks, each with a unique identifier. Block storage allows the underlying storage system to retrieve it regardless of where it is stored.

File storage is the most well-known storage format: data is stored in files that can be interacted with, contained in folders within a hierarchical file directory.

Object storage is a storage format in which data is stored in separate units called objects. Each unit has a unique identifier, or key, that allows it to be located independently of where it is stored in a distributed system.

File Storage

Unix and Unix Like

- **Name**
- Path
- Type
- Size
- Owner (UID, GID)
- Permission
- Timestamps
 - creation
 - modify
- All file names are "Case Sensitive." This means that vivek.txt, Vivek.txt, and VIVEK.txt are three different files.
- File names can use uppercase and lowercase letters as well as the symbols "." (dot) and "_" (underscore).
- Other special characters like " " (blank space) can also be used, but they require a more complex handling (they must be quoted) and are generally discouraged.
- In practice, a file name can contain any character except "/" (root folder), which is reserved as a separator between files and folders in the pathname.
- The ***null*** character cannot be used.
- Using a "." is not necessary but increases readability, especially if used to identify the file extension.
- The file name must be unique within a folder.
- A folder and a file with the same name cannot coexist within the same folder.

Maxlength 255 characters

File Storage

Unix and Unix Like

- **Name**
- Path
- Type
- Size
- Owner (UID, GID)
- Permission
- Timestamps
 - creation
 - modify

Il set di nomi richiesto per specificare un particolare file in una gerarchia di folder è detto percorso del file o path.
percorso e nome del file formano il cosiddetto pathname.

Il percorso può essere assoluto o relativo:

- nel path assoluto si specifica tutto il percorso dall'inizio del disco (/root):
`/u/politi/projectb/plans/1dft`
- nel path relativo si può indicare il percorso a partire dal folder in cui ci si trova.
`projectb/plans/1dft`

Un path relativo non può iniziare con /.

Simboli speciali:

- indica il folder corrente
- .. indica il folder di livello superiore

File Storage

Unix and Unix Like

- Name
- **Path**
- Type
- Size
- Owner (UID, GID)
- Permission
- Timestamps
 - creation
 - modify

The set of names required to specify a particular file in a hierarchy of folders is called the file path.

The path and the filename together form what is known as the pathname.

The path can be absolute or relative:

- In an absolute path, the entire path is specified starting from the beginning of the disk (/ , root):
/u/politi/projectb/plans/ldft
- In a relative path, the path can be indicated starting from the folder in which you are located:
projectb/plans/ldft

A relative path cannot start with /.

Special symbols:

- . indicates the current folder
- .. indicates the parent folder

Maxlength 1024 characters

File Storage

Unix and Unix Like

- Name
- Path
- Type
- Size
- Owner (UID, GID)
- Permission
- Timestamps
 - creation
 - modify

The type of file is identified by the first character of the permission string.

```
-rwxrwxrwx 1 romolo romolo      658 apr 30 09:56 manage.py
```

The types could be:

-	regular file
d	directory
l	symbolic link
c	Character file device
b	block device
s	local socket
p	named pipe

Lib.

- like

The types could be:

- regular file
- d directory
- l symbolic link
- c Character file device
- b block device
- s local socket
- p named pipe

```
658 apr 30 09:56 manage.py
```


Lib.

- Like

```
-rwxrwxrwx 1 romolo romolo 658 apr 30 09:56 manage.py
```

The types could be:

[https://it.wikipedia.org/wiki/Permessi_\(informatica\)](https://it.wikipedia.org/wiki/Permessi_(informatica))

File Storage

Unix and Unix Like

- Name
- Path
- **Type**
- Size
- Owner (UID, GID)
- Permission
- Timestamps
 - creation
 - modify

The type of file is identified by the first character of the permission string.

```
-rwxrwxrwx 1 romolo romolo      658 apr 30 09:56 manage.py
```

The types could be:

-	regular file
d	directory
l	symbolic link
c	Character file device
b	block device
s	local socket
p	named pipe

Object Storage

In object storage, data is broken down into discrete units called objects and stored in a single repository (**Bucket**) instead of as files within folders or as blocks on servers.

The volumes of object storage function as modular units: each one is an independent repository that contains the data, a unique identifier that allows an object to be located in a distributed system, and the metadata that describes the data.

Metadata is important and includes details such as age, privacy/security, and access restrictions.

Object Storage

In object storage, data is broken down into discrete units called objects and stored in a single repository (**Bucket**) instead of as files within folders or as blocks on servers.

The volume
repository
in a distrib

Object storage **metadata** can be extremely detailed and capable of storing information about where a video was filmed, the type of camera used, and the actors appearing in each frame.

pendent
e located

Metadata is important and includes details such as age, privacy/security, and access restrictions.

The Data Preservation

In data management, **Data Preservation** is the act of safeguarding and maintaining both the security and integrity of data. Preservation is achieved through formal activities governed by policies, regulations, and strategies designed to protect and prolong the existence and authenticity of data and its associated metadata.

The Data Preservation

In data management, **Data Preservation** is the act of safeguarding and maintaining both the security and integrity of data. Preservation is achieved through formal activities governed by policies, regulations, and strategies designed to protect and prolong the existence and authenticity of data and its associated metadata.

- short-term
- medium-term
- long-term

The Data Preservation

In data management, **Data Preservation** is the act of safeguarding and maintaining both the security and integrity of data. Preservation is achieved through formal activities governed by policies, regulations, and strategies designed to protect and prolong the existence and authenticity of data and its associated metadata.

- short-term
- ~~medium-term~~
- long-term

The Data Preservation

In data management, **Data Preservation** is the act of safeguarding and maintaining both the security and integrity of data. Preservation is achieved through formal activities governed by policies, regulations, and strategies designed to protect and prolong the existence and authenticity of data and its associated metadata.

- **short-term**
- ~~medium-term~~
- long-term

Short-term Preservation.

Access to digital materials for a defined period during which use is expected, but which does not extend beyond the foreseeable future and/or until it becomes inaccessible due to technological changes.

The Data Preservation

In data management, **Data Preservation** is the act of safeguarding and maintaining both the security and integrity of data. Preservation is achieved through formal activities governed by policies, regulations, and strategies designed to protect and prolong the existence and authenticity of data and its associated metadata.

- short-term
- ~~medium-term~~
- long-term

Long-term Preservation

Continuous access to digital materials, or at least to the information contained in them, indefinitely.

Version Control



Git

Git is a distributed version control software that can be used via the command line interface, created by Linus Torvalds in 2005.

A **Distributed Version Control System** (DVCS) is a type of version control that allows tracking changes and versions made to software source code without needing to use a central server, as in traditional cases.

With this system, developers can collaborate individually and in parallel, working on their own development **branch**, recording their changes (**commits**), and later sharing or **merging** them with others' changes—all without the need for a centralized server. This system enables various modes of collaboration, as the server is merely a support tool.



Glossary

repository: It is a "folder" that contains all the files needed for your project, including the files that track all versions of the project.

clone: It is the local version of the repository.

remote: It is the remote version of the repository that can be modified by anyone with access to the repository.

branch: "Branches" are used in Git to implement isolated features, that is, developed independently from each other but starting from the same root.

fork: A copy of a repository belonging to another user.

commit: A snapshot of the local repository compressed with SHA, ready to be transferred from the clone to the remote or vice versa.

tag: A marker used to highlight specific commits.



First Steps

Git can be downloaded from <https://git-scm.com/downloads> (all Linux distributions have Git among the available packages).

Once the software is installed, to "copy" a repository locally, you simply use the clone command.

For example, for the course repository:

```
git clone https://github.com/RomoloPoliti-INAF/PhDCourse2024.git
```



First steps

Git can be downloaded from <https://git-scm.com/downloads> (all Linux distributions have Git among the available packages).

Once the software is installed, to "copy" a repository locally, you simply use the clone command.

For example, for the course repository:

```
git clone https://github.com/RomoloPoliti-INAF/PhDCourse2024.git
```

Windows users need to install the Git application by downloading it from the [GitHub](https://github.com) website.
After installation, it will be necessary to restart the machine.



Fundamental Git commands

clone: Create a local copy of a remote repository

pull: Update the local copy of the repository

add: Add one or more files to the list of contents in the local repository

commit: Record changes to the repository

push: Update the remote repository