

Cloud Archive and Computation

...

Romolo Politi

Elenco Lezioni

Elenco Lezioni

- 01/06/2023

Panoramica del Corso

Panoramica del Corso

Cloud

Struttura del Cloud
Dati nel Cloud
Calcolo nel Cloud

Dati

Dati e metadati
Archiviazione
DB Relazionali e non

Calcolo

Recupero
Manipolazione
Visualizzazione

Ambiente

Virtualizzazione e container
Microservices
DevOps

Programmazione

Fondamenti di programmazione
Python
Versioning e Documentazione

Tools

- La presentazione e gli esempi del corso sono su GitHub:
 - <https://github.com/RomoloPoliti-INAF/PhDCourse2023>
- Per gli esempi utilizzeremo Python 3.11.1
- Come framework di sviluppo Microsoft Visual Studio Code
 - <https://code.visualstudio.com>

Struttura del Corso

- La lista degli argomenti mostrata in precedenza è stata costruita per categorie.
- Noi seguiremo un percorso guidato dagli esempi per meglio capire la filosofia che c'è dietro.
- Dopo l'introduzione alla programmazione svilupperemo un esempio di programma complesso (Macchina a Stati).
- In ultimo svilupperemo una WebApp e la prepareremo per la distribuzione su container
- Per alcuni argomenti non scenderemo nel dettaglio perché lo scopo del corso è dare una visione generale della tematica.
- Anche se non verranno discussi, molti dettagli saranno nelle slide o nei link riportati.

Lezione del 01/06/2023

Definizione di Cloud

Cosa è il Cloud

E' una erogazione di servizi offerti su richiesta da un fornitore ad un utente finale attraverso la rete internet (come l'archiviazione, l'elaborazione o la trasmissione dati), a partire da un insieme di risorse preesistenti, configurabili e disponibili in remoto sotto forma di architettura distribuita.

wikipedia

Tipi di Cloud

In Promise



Out Promise



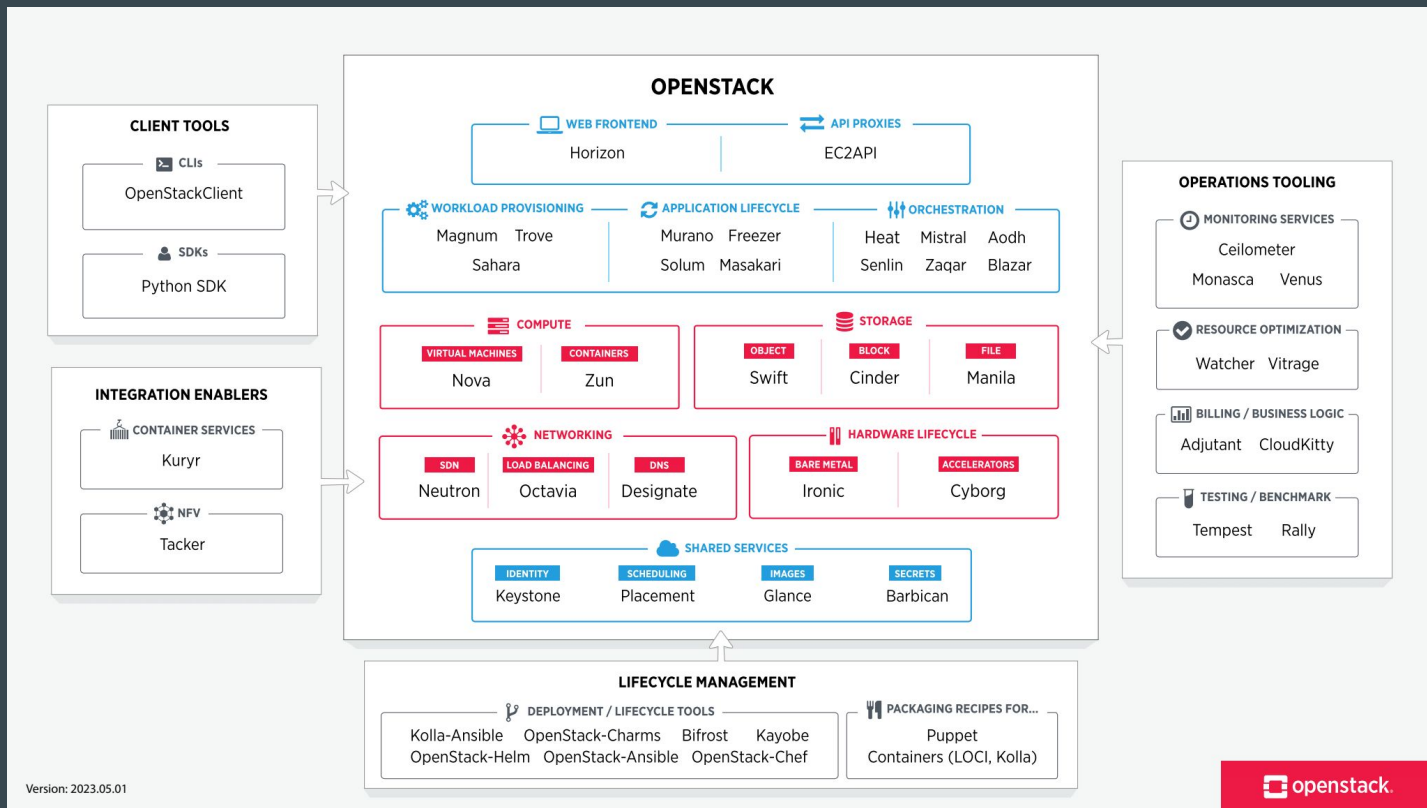
Google Cloud



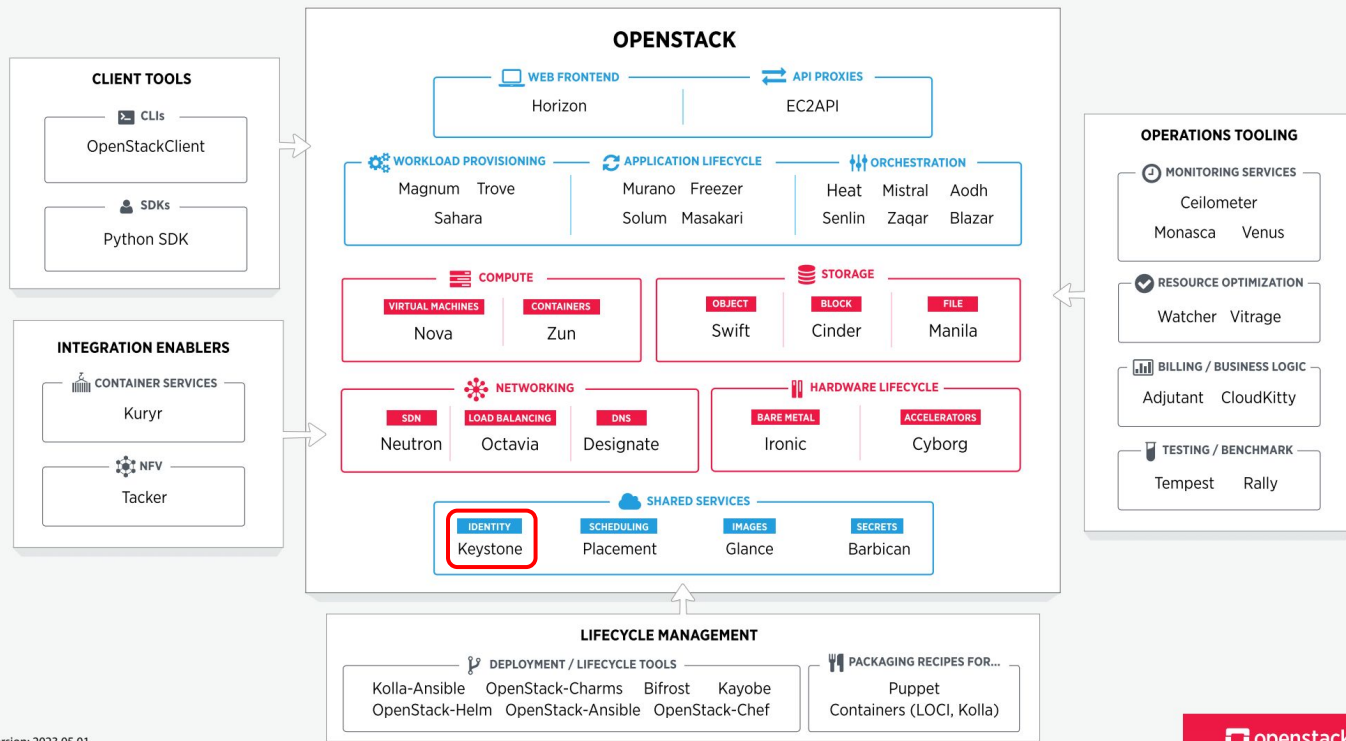
Struttura del Cloud

Struttura di un Cloud

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

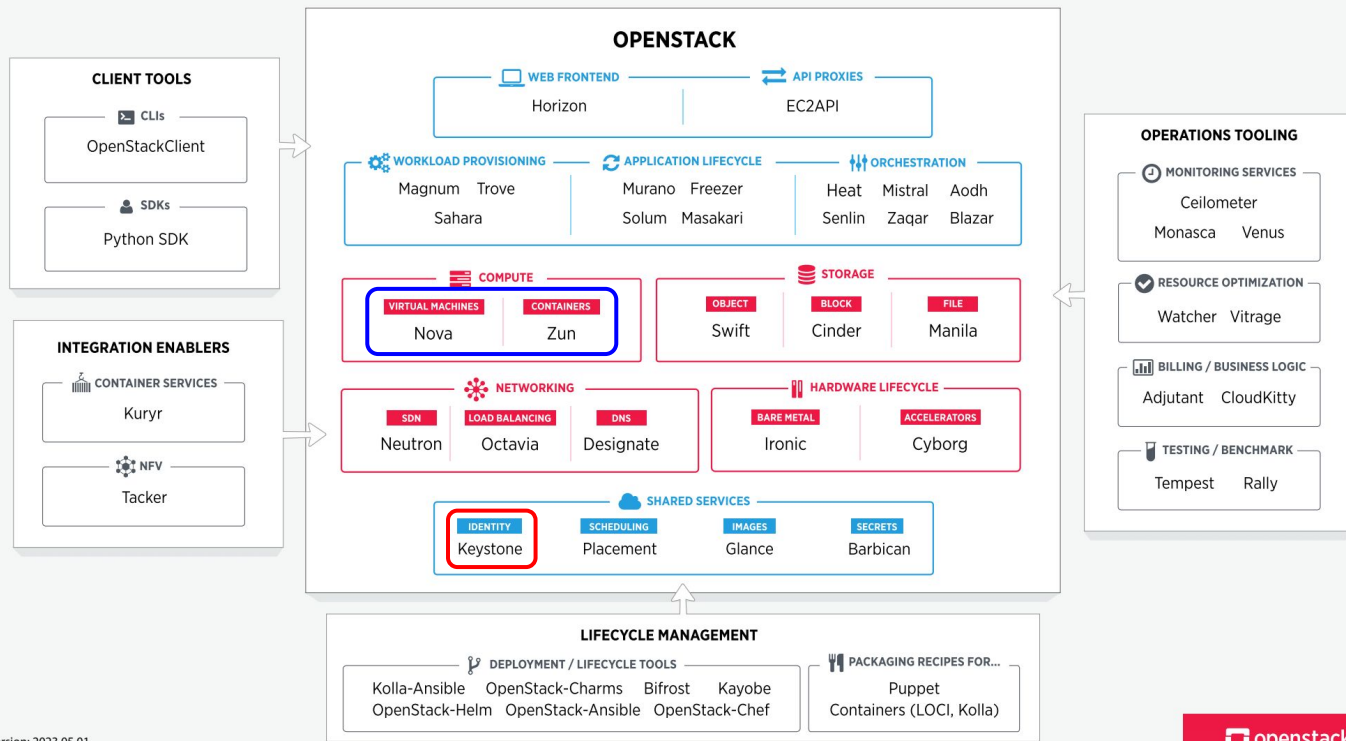


Struttura di un Cloud



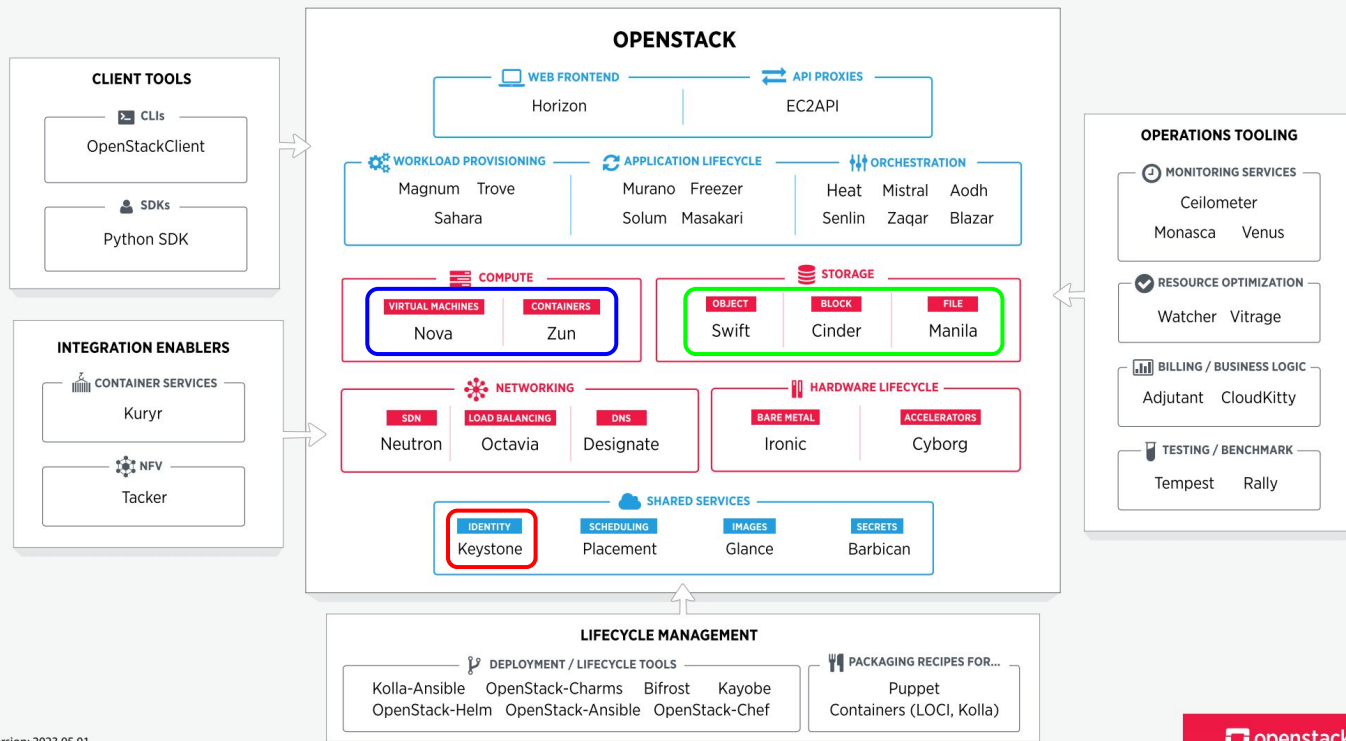
- Identity

Struttura di un Cloud



- Identity
- Compute

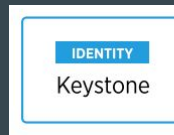
Struttura di un Cloud



- Identity
- Compute
- Storage

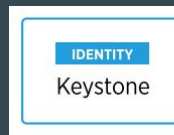
Componenti Principali

- IAM (Identity and Access Management)



Componenti Principali

- IAM (Identity and Access Management)
 - verifica identità
 - lista di risorse dedicate
 - privilegi
 - Credito (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 labeled "Stay signed in", and at the bottom right is a blue link labeled "Need help?".

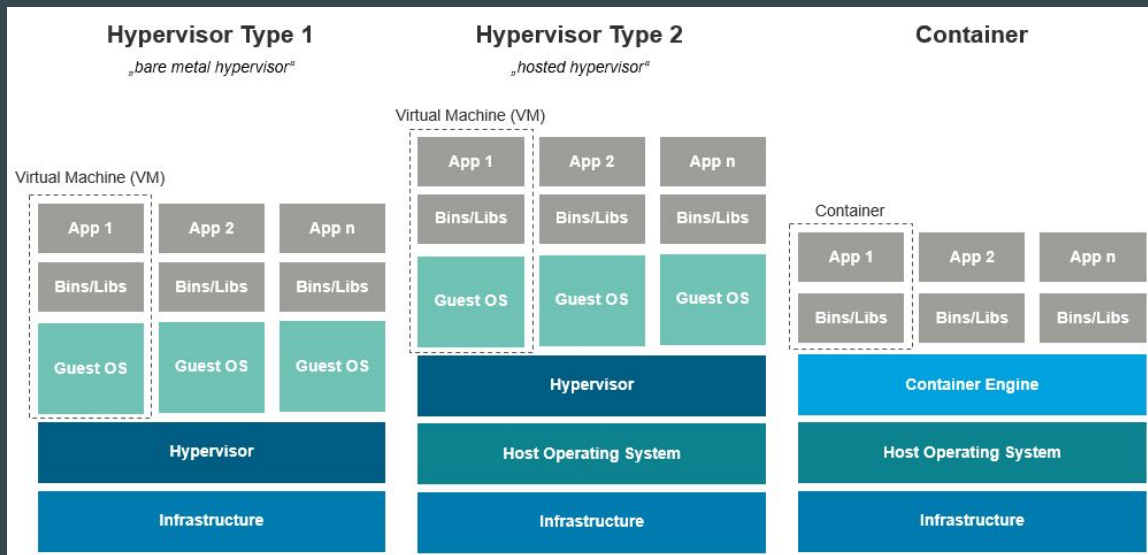
Componenti Principali

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



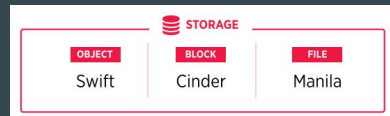
Componenti Principali

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



Componenti Principali

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

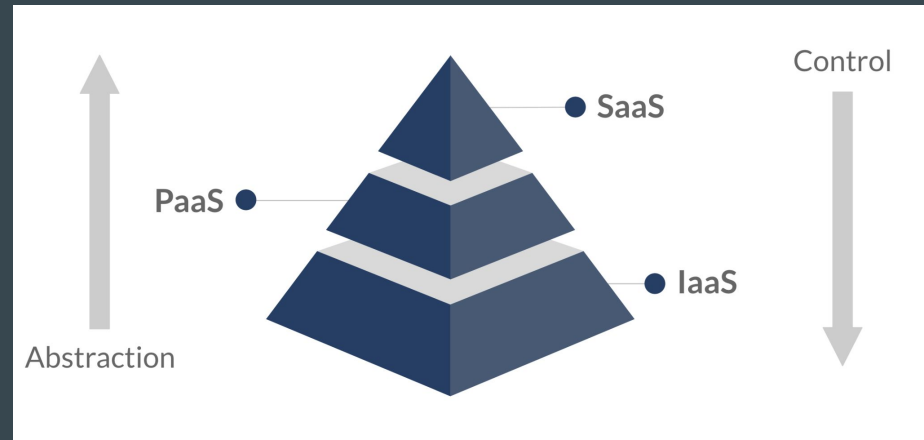


Servizi Cloud

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

Servizi Cloud

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

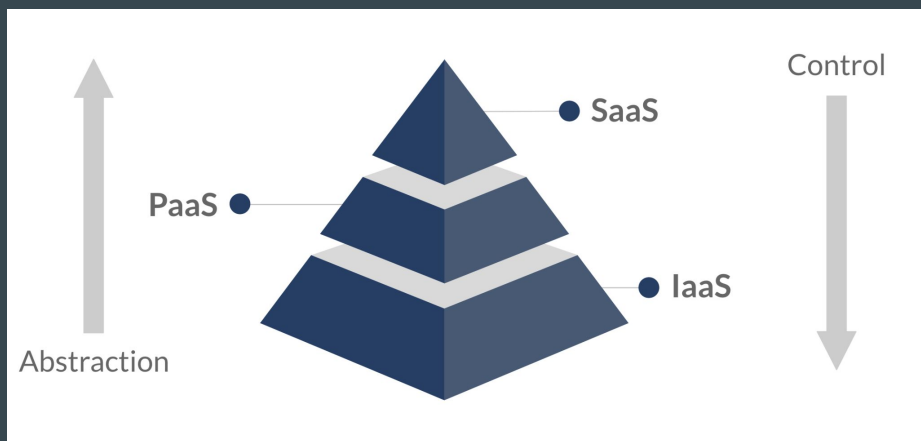


IaaS

Il provider offre un hardware virtuale (CPU, RAM, spazio e schede di rete) e quindi la flessibilità di un'infrastruttura fisica, senza l'onere per l'utente, della gestione fisica dell'hardware

Servizi Cloud

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

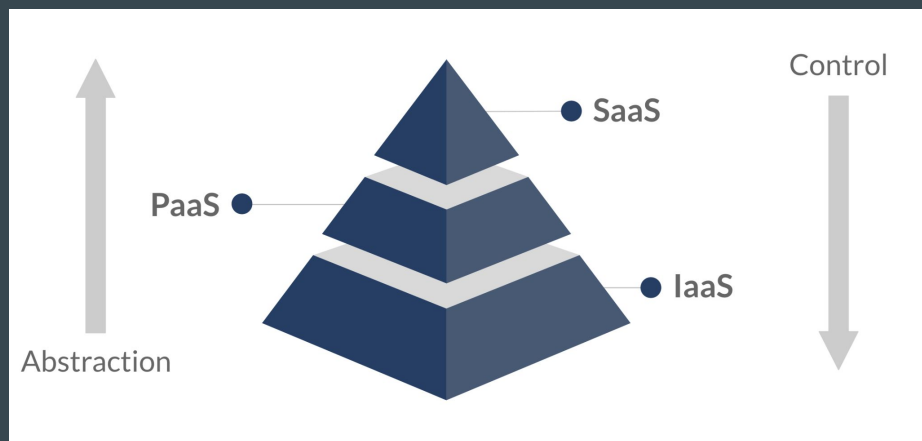


PaaS

Il provider si occupa dell'infrastruttura hardware, mentre l'utente dovrà installare il sistema operativo e occuparsi di sviluppare la sua applicazione

Servizi Cloud

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

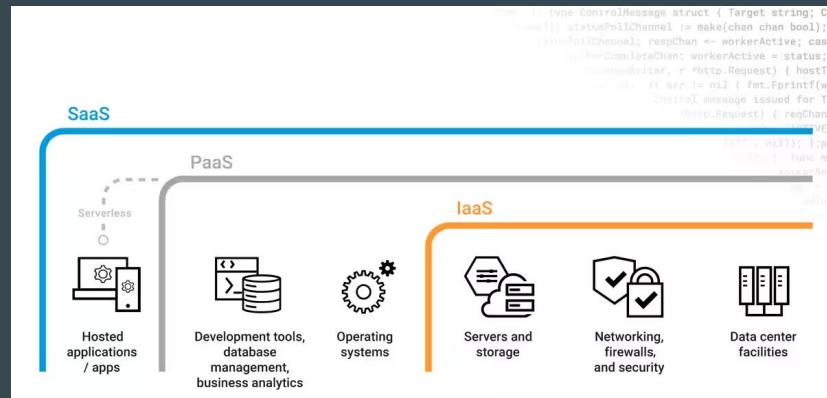
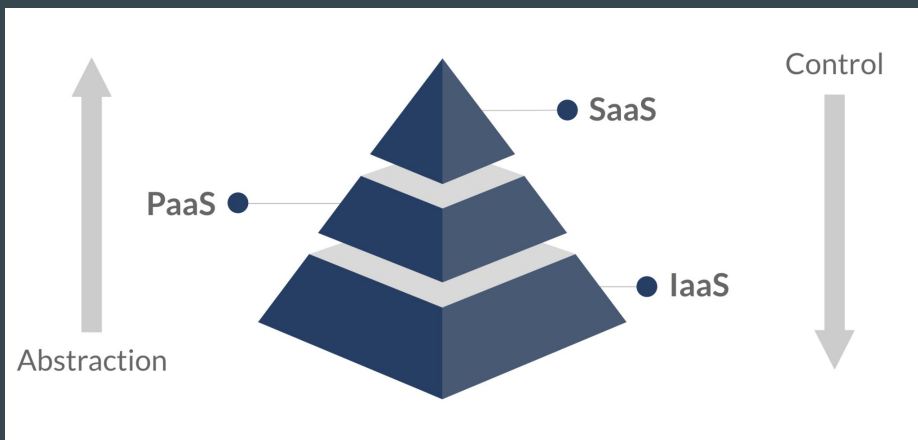


SaaS

L'utente finale non ha bisogno di nessuna conoscenza informatica per utilizzare l'applicazione o i servizi erogati. I servizi sono utilizzabili semplicemente con una connessione internet e un browser.

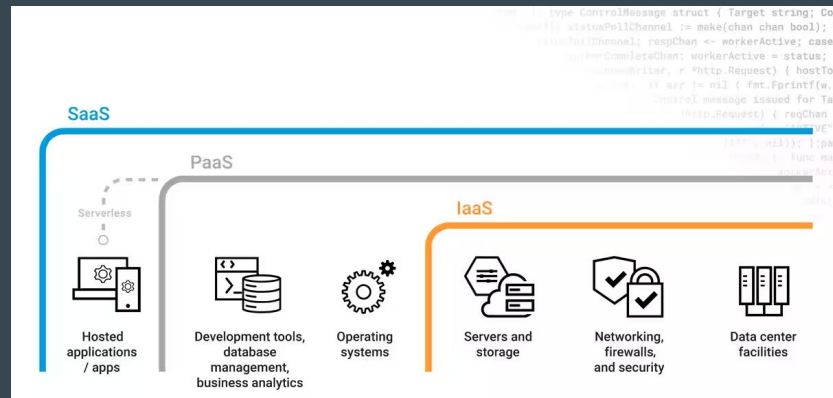
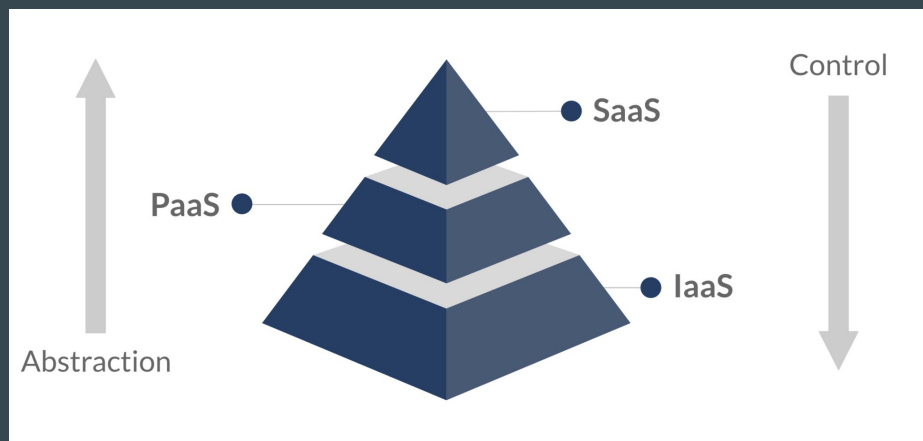
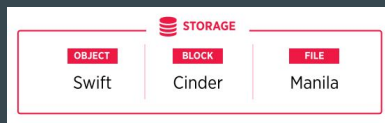
Servizi Cloud

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



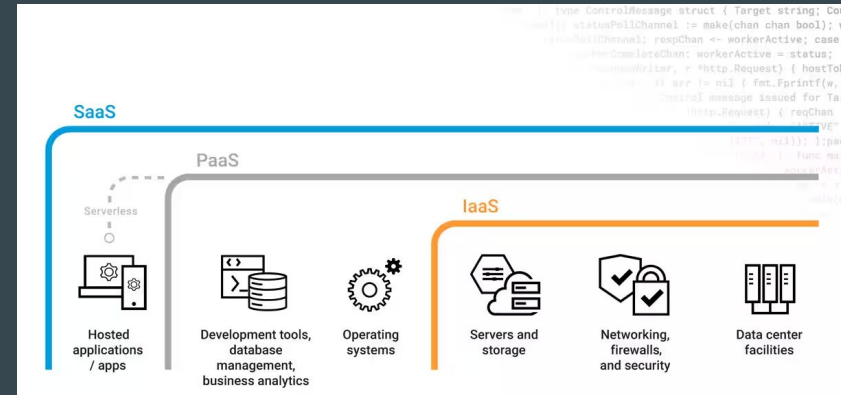
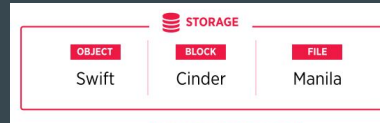
Servizi Cloud

- 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



Dati e Metadati

Definizione di Dato

Un dato (dal latino *datum* dono, cosa data) è una descrizione elementare codificata di un'informazione, un'entità, di un fenomeno, di una transazione, di un avvenimento o di altro.

Un dato (in informatica) può avere dimensione da 1 bit (booleano) sino a migliaia di milioni di byte.

Definizione di Metadato

Il metadato è, letteralmente, "(dato) per mezzo di un (altro) dato", è un'informazione che descrive un insieme di dati.

Un esempio tipico di metadati è costituito dalla scheda del catalogo di una biblioteca, la quale contiene informazioni circa il contenuto e la posizione di un libro, cioè dati riguardanti più dati che si riferiscono al libro. Un altro contenuto tipico dei metadati può essere la fonte o l'autore dell'insieme di dati descritto, oppure le modalità d'accesso con le eventuali limitazioni.

Un metadato può essere anche un dato aggiunto all'insieme delle informazioni per altri scopi. Ad esempio, se alla scheda del libro della biblioteca aggiungo un ID, ossia un identificatore univoco, quest'ultimo è un metadato.

Esempio Dato - Metadato



Informazioni

Aggiungi una descrizione

DETTAGLI

20 lug 2019
sab, 13:48 GMT+02:00

motorola moto g(6)
f/1.8 1/890 3,95 mm ISO 100

IMG_20190720_134639156_HDR.jpg
12,6 MP 4096 x 3072

Caricata da un dispositivo Android

Backup eseguito
Risparmio spazio di archiviazione. Scopri di più

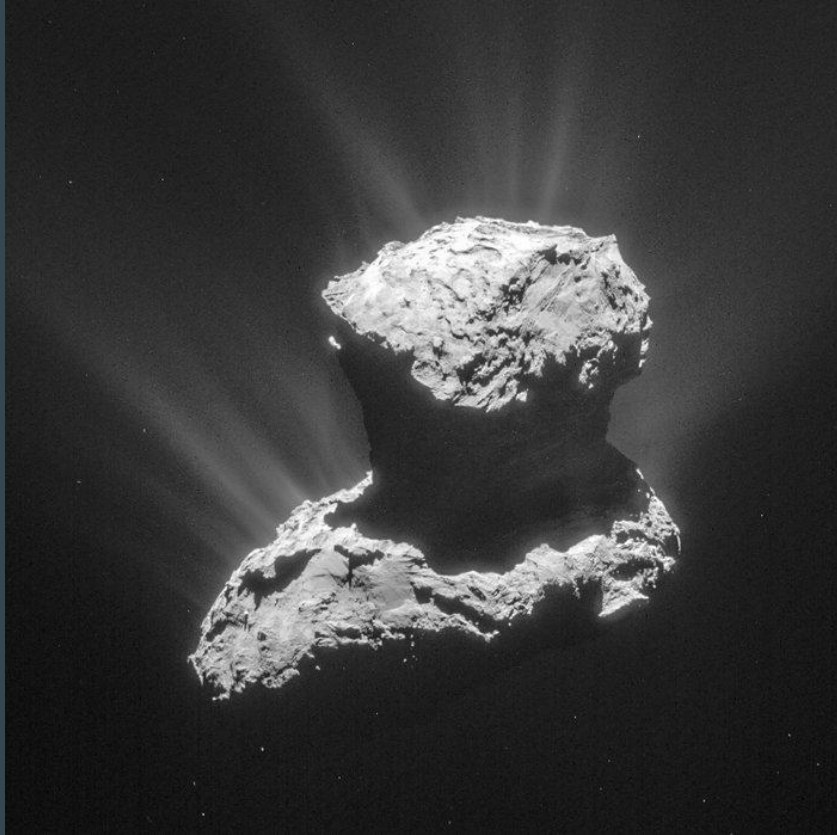
Questo elemento non occupa spazio di archiviazione dell'account. Scopri di più

Amalfi Provincia di Salerno

Pontone
Terrazza dell'Infinito
Museo della Carta
Lido di Ravello
Spargosa di Caricigione
POGEROLA
Atrani
Amalfi
Santa Capriola

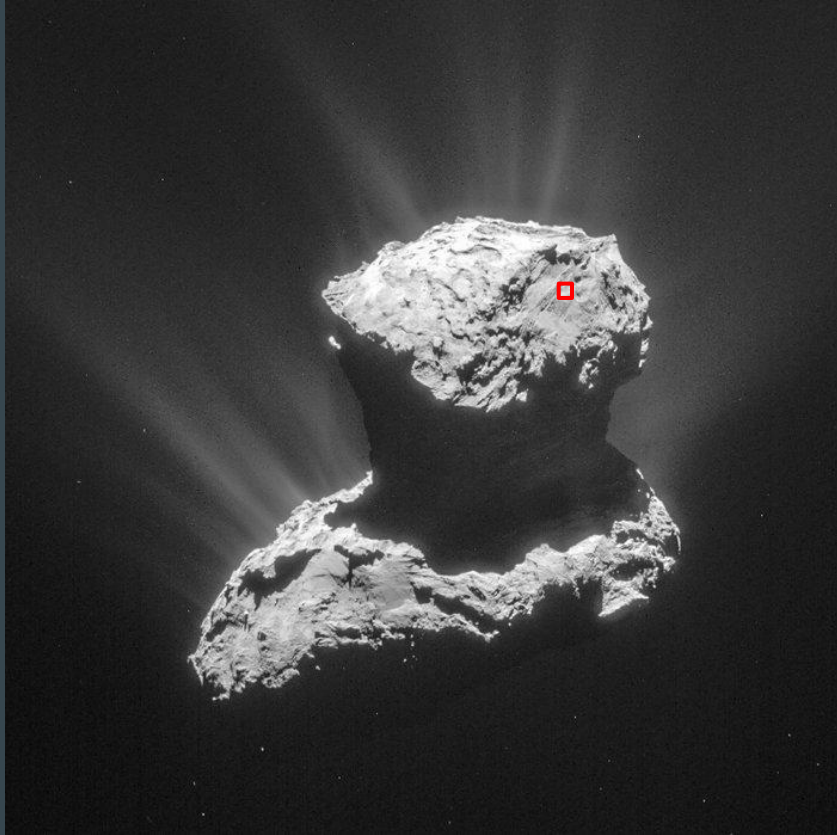
Map data ©2023

Esempio in Planetologia



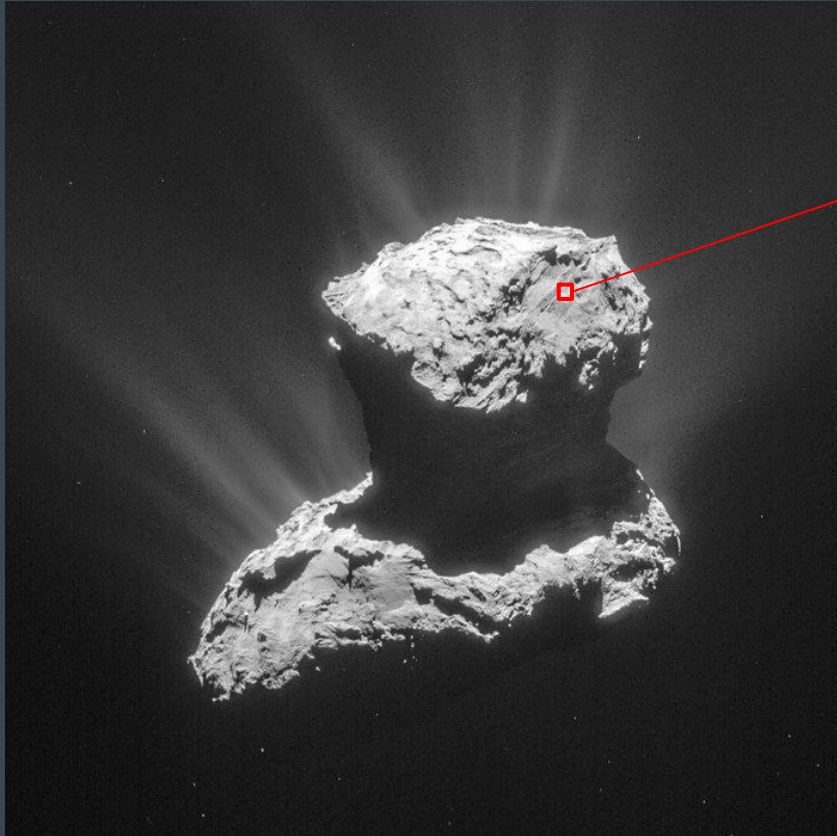
Quale è il dato in questo caso?

Esempio in Planetologia



Quale è il dato in questo caso?

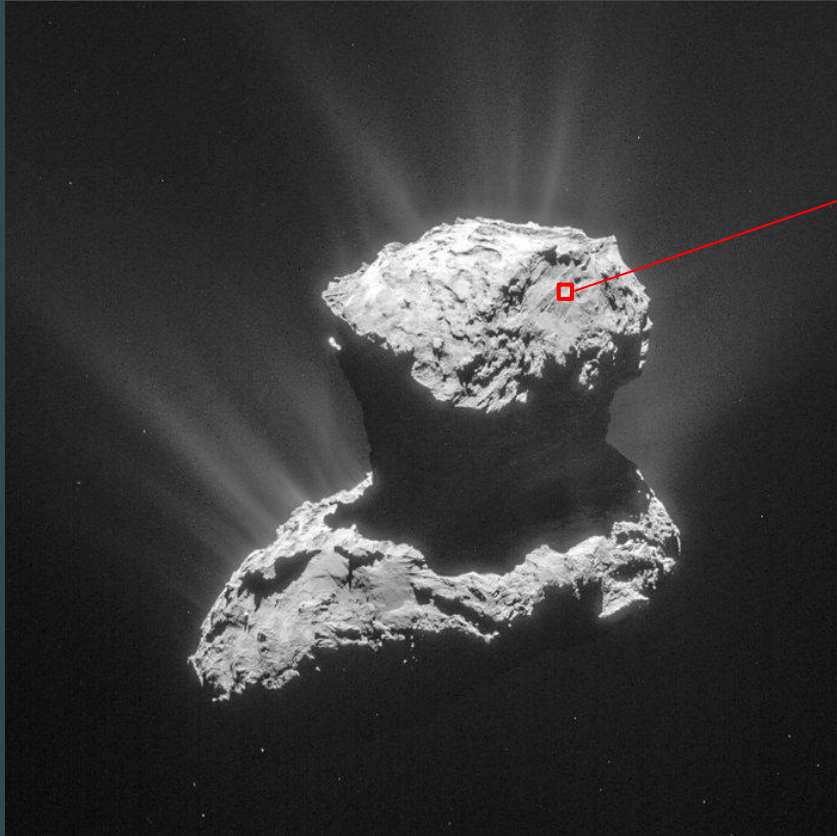
Esempio in Planetologia



Quale è il dato in questo caso?

Il pixel è rappresentato come un numero in virgola mobile a 32 bit.

Esempio in Planetologia

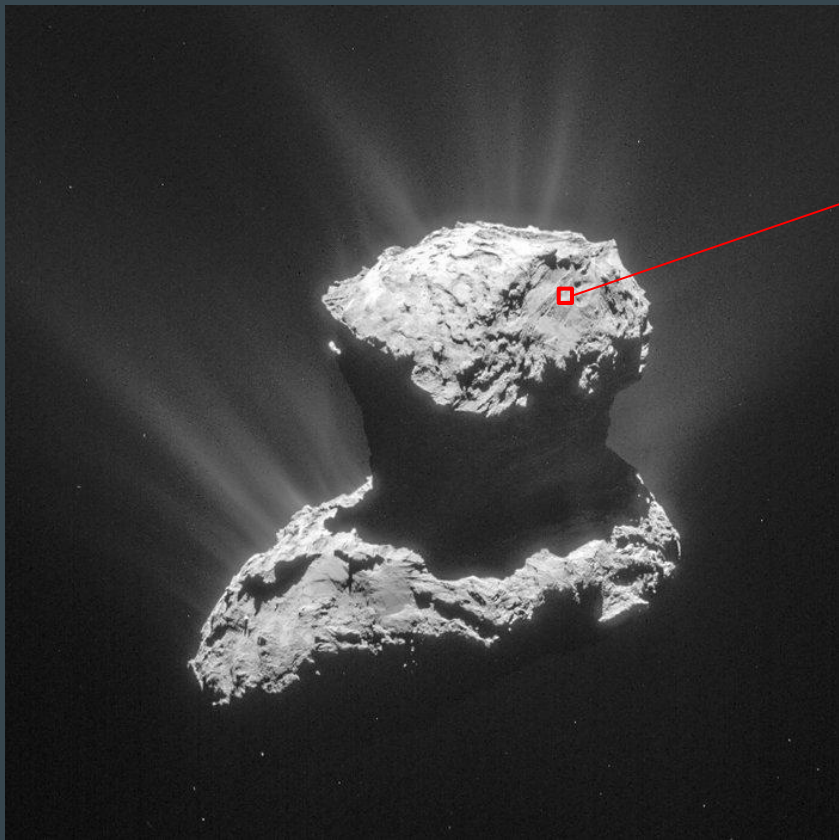


Quale è il dato in questo caso?

Il pixel è rappresentato come un numero in virgola mobile a 32 bit.

Ha significato?

Esempio in Planetologia



Quale è il dato in questo caso?

Il pixel è rappresentato come un numero in virgola mobile a 32 bit.

Ha significato?

La risposta è **NO**.

Si ha necessità di conoscere

- illuminazione,
- posizione della cometa,
- posizione dello spacecraft,
- tempi di esposizione,
- modalità di acquisizione,
- georeferenziazione del pixel

Dati nel Cloud

Archiviazione nel Cloud

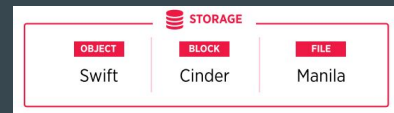


Archiviazione nel Cloud



Il **block storage** suddivide i dati in componenti separati composti da blocchi di dati di dimensioni fisse, ognuno dotato di un identificatore univoco. Il block storage permette al sistema di storage sottostante di recuperarlo indipendentemente dalla posizione in cui viene memorizzato.

Archiviazione nel Cloud



Il **block storage** suddivide i dati in componenti separati composti da blocchi di dati di dimensioni fisse, ognuno dotato di un identificatore univoco. Il block storage permette al sistema di storage sottostante di recuperarlo indipendentemente dalla posizione in cui viene memorizzato.

Il **file storage** è il formato di storage maggiormente conosciuto: i dati vengono archiviati in file con cui è possibile interagire, contenuti in cartelle all'interno di una directory file gerarchica.

Archiviazione nel Cloud



Il **block storage** suddivide i dati in componenti separati composti da blocchi di dati di dimensioni fisse, ognuno dotato di un identificatore univoco. Il block storage permette al sistema di storage sottostante di recuperarlo indipendentemente dalla posizione in cui viene memorizzato.

Il file storage è il formato di storage maggiormente conosciuto: i dati vengono archiviati in file con cui è possibile interagire, contenuti in cartelle all'interno di una directory file gerarchica.

L'**object storage** è un formato di storage in cui i dati sono archiviati in unità separate chiamate oggetti. Ciascuna unità ha un identificatore univoco, o chiave, che ne permette l'individuazione indipendentemente dalla posizione in cui sono memorizzate in un sistema distribuito.

Archiviazione nel Cloud



Il **block storage** suddivide i dati in componenti separati composti da blocchi di dati di dimensioni fisse, ognuno dotato di un identificatore univoco. Il block storage permette al sistema di storage sottostante di recuperarlo indipendentemente dalla posizione in cui viene memorizzato.

Il **file storage** è il formato di storage maggiormente conosciuto: i dati vengono archiviati in file con cui è possibile interagire, contenuti in cartelle all'interno di una directory file gerarchica.

L'**object storage** è un formato di storage in cui i dati sono archiviati in unità separate chiamate oggetti. Ciascuna unità ha un identificatore univoco, o chiave, che ne permette l'individuazione indipendentemente dalla posizione in cui sono memorizzate in un sistema distribuito.

File Storage

Unix and Unix Like

- **Nome**
- Percorso (path)
- Tipo
- Dimensione
- Proprietario (UID, GID)
- Permessi
- Marcature Temporali
 - creazione
 - modifica
- Tutti i nomi dei file sono “Case Sensitive”. Ciò vuol dire che vivek.txt Vivek.txt VIVEK.txt sono tre file differenti.
- Per i nomi di file si possono usare lettere maiuscole, minuscole ed i simboli “.” (dot), e “_” (underscore).
- Possono essere usati anche altri caratteri speciali come “ ” (blank space) ma hanno un uso complesso (devono essere quotati) e se ne sconsiglia l’uso.
- In pratica il nome di un file può contenere qualsiasi carattere escluso “/” (root folder) che è riservato come separatore tra file e folder nel pathname.
- Non può essere usato il carattere null.
- L’uso del “.” non è necessario ma aumenta la leggibilità specialmente se usato per identificare l’estensione.
- Il nome del file è unico all’interno di un folder.
- In un folder non possono coesistere un folder ed un file con lo stesso nome.

Lunghezza massima 255 caratteri

File Storage

Unix and Unix Like

- Nome
- **Percorso (path)**
- Tipo
- Dimensione
- Proprietario (UID, GID)
- Permessi
- Marcature Temporali
 - creazione
 - modifica

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

- Nome
- **Percorso (path)**
- Tipo
- Dimensione
- Proprietario (UID, GID)
- Permessi
- Marcature Temporali
 - creazione
 - modifica

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

- Nome
- Percorso (path)
- **Tipo**
- Dimensione
- Proprietario (UID, GID)
- Permessi
- Marcature Temporali
 - creazione
 - modifica

Il tipo di file viene identificato dal primo carattere della stringa dei permessi.

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

I tipi possono essere:

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

Lib.

- like

annunciato da

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

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

Like

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

- file regolare

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

# File Storage

## Unix and Unix Like

- Nome
- Percorso (path)
- Tipo
- **Dimensione**
- Proprietario (UID, GID)
- Permessi
- Marcature Temporali
  - creazione
  - modifica

Il tipo di file viene identificato dal primo carattere della stringa dei permessi.

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

I tipi possono essere:

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

# Object Storage

Nello storage di oggetti, i dati vengono frammentati in unità discrete chiamate appunto oggetti e conservati in un unico repository (Bucket) invece che come file all'interno di cartelle o come blocchi su server.

I volumi dello storage di oggetti operano come unità modulari: ognuno è un repository indipendente che conserva al suo interno i dati, un identificativo univoco che permette di individuare un oggetto in un sistema distribuito e i metadati che descrivono i dati. I metadati sono importanti e includono dettagli come l'età, privacy/sicurezza e limitazioni all'accesso.

# Object Storage

Nello storage di oggetti, i dati vengono frammentati in unità discrete chiamate appunto oggetti e conservati in un unico repository (Bucket) invece che come file all'interno di cartelle o come documenti.

I volumi dei dati sono indipendenti

I **metadati** dello storage di oggetti possono essere estremamente dettagliati e capaci di archiviare informazioni sul luogo in cui un video è stato girato, sul tipo di fotocamera che è stato utilizzato e sugli attori che compaiono in ogni fotogramma.

repository  
permette

di individuare un oggetto in un sistema distribuito e i metadati che descrivono i dati. I metadati sono importanti e includono dettagli come l'età, privacy/sicurezza e limitazioni all'accesso.

# Concetto di Data Preservation

Nella gestione dei dati (Data Management) la **Data Preservation** è l'atto di conservare e mantenere sia la sicurezza che l'integrità dei dati. La conservazione avviene attraverso attività formali disciplinate da politiche, regolamenti e strategie volte a proteggere e prolungare l'esistenza e l'autenticità dei dati e dei relativi metadati.

# Concetto di Data Preservation

Nella gestione dei dati (Data Management) la **Data Preservation** è l'atto di conservare e mantenere sia la sicurezza che l'integrità dei dati. La conservazione avviene attraverso attività formali disciplinate da politiche, regolamenti e strategie volte a proteggere e prolungare l'esistenza e l'autenticità dei dati e dei relativi metadati.

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

# Concetto di Data Preservation

Nella gestione dei dati (Data Management) la **Data Preservation** è l'atto di conservare e mantenere sia la sicurezza che l'integrità dei dati. La conservazione avviene attraverso attività formali disciplinate da politiche, regolamenti e strategie volte a proteggere e prolungare l'esistenza e l'autenticità dei dati e dei relativi metadati.

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

# Concetto di Data Preservation

Nella gestione dei dati (Data Management) la **Data Preservation** è l'atto di conservare e mantenere sia la sicurezza che l'integrità dei dati. La conservazione avviene attraverso attività formali disciplinate da politiche, regolamenti e strategie volte a proteggere e prolungare l'esistenza e l'autenticità dei dati e dei relativi metadati.

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

## **Conservazione a breve termine.**

Accesso ai materiali digitali per un periodo di tempo definito durante il quale è previsto l'uso ma che non si estende oltre il prevedibile futuro e/o fino a quando non diventa inaccessibile a causa dei cambiamenti tecnologici.



# Concetto di Data Preservation

Nella gestione dei dati (Data Management) la **Data Preservation** è l'atto di conservare e mantenere sia la sicurezza che l'integrità dei dati. La conservazione avviene attraverso attività formali disciplinate da politiche, regolamenti e strategie volte a proteggere e prolungare l'esistenza e l'autenticità dei dati e dei relativi metadati.

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

## Conservazione a lungo termine

Accesso continuo ai materiali digitali, o almeno alle informazioni in essi contenute, a tempo indeterminato.

# Controllo di versione

# Git



**Git** è un software per il controllo di versione distribuito utilizzabile da interfaccia a riga di comando, creato da Linus Torvalds nel 2005.

Un sistema di controllo di versione distribuito o decentralizzato (o **DVCS** da Distributed Version Control System) è una tipologia di controllo di versione che permette di tenere traccia delle modifiche e delle versioni apportate al codice sorgente del software, senza la necessità di dover utilizzare un server centrale, come nei casi classici.

Con questo sistema gli sviluppatori possono collaborare individualmente e parallelamente non connessi su di un proprio ramo (branch) di sviluppo, registrare le proprie modifiche (commit) ed in seguito condividerle con altri o unirle (merge) a quelle di altri, il tutto senza bisogno del supporto di un server centralizzato. Questo sistema permette diverse modalità di collaborazione, proprio perché il server è soltanto un mero strumento d'appoggio.



# Glossario

- repository:** È una “cartella” che contiene tutti i file necessari per il tuo progetto, inclusi i file che tengono traccia di tutte le versioni del progetto.
- clone:** È la versione locale del repository
- remote:** È la versione remota del repository che può essere modificata da chiunque abbia accesso al repository.
- branch:** “rami” vengono utilizzati in Git per l'implementazione di funzionalità tra loro isolate, cioè sviluppate in modo indipendente l'una dall'altra ma a partire dalla medesima radice.
- fork:** copia del repository appartenente ad un altro utente
- commit:** snapshot del repository locale compresso con SHA pronto per essere trasferito, dal clone al remote o viceversa.
- tag:** è un marcatore per evidenziare dei particolari commit



# Primi Passi

Git può essere scaricato all'indirizzo <https://git-scm.com/downloads> (tutte le distribuzioni di linux hanno git tra i pacchetti disponibili).

Una volta installato il software, per “copiare” un repository in locale basta utilizzare il comando clone. Ad esempio per il repository del corso:

```
git clone git@github.com:RomoloPoliti-INAF/PhDCourse2023.git
```



# Comandi Git fondamentali

- clone** Crea una copia locale di un repository remoto
- pull** Aggiorna la copia locale del repository
- add** Aggiungi uno o più file alla lista dei contenuti del repository locale
- commit** Registra i cambiamenti al repository
- push** aggiorna il repository remoto