Crittografia per la privacy nei big-data

Marco **Pedicini**
Dipartimento di Matematica e Fisica
Università Roma Tre

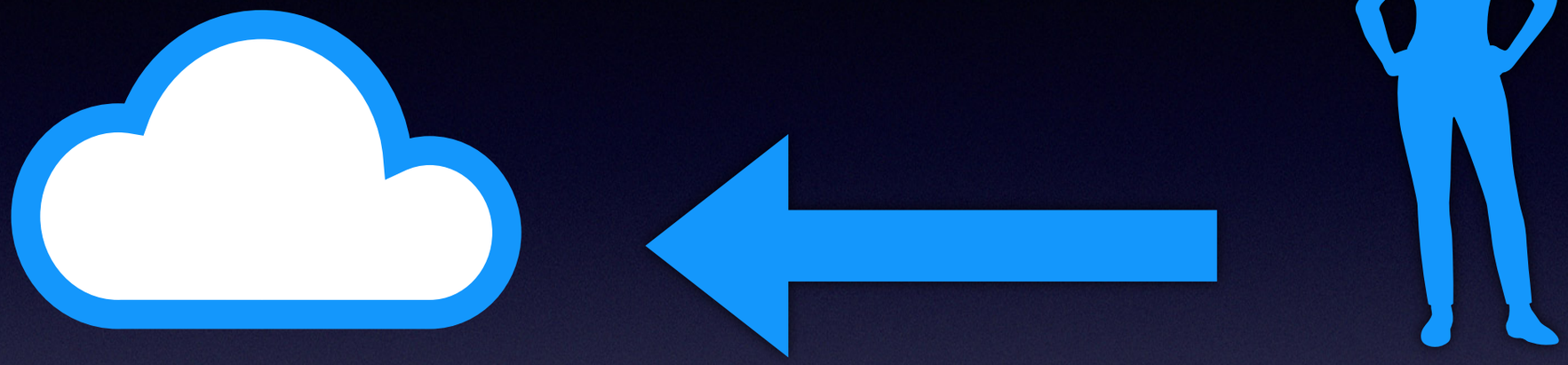
Data out-sourcing

- Big Data per molte organizzazioni significa combattere con una **enorme quantità di dati**, tanto da essere costretti a conferire tali **dati a terze parti**.
- Al fine di proteggere questi dati, le **basi di dati esterne** vengono protette tramite varie tecniche crittografiche, che oltre a garantite la **privacy** devono permettere l'**accesso efficiente** ai dati.
- Recentemente abbiamo assistito a vari **attacchi** a sistemi basati su questi principi — soprattutto a causa del **compromesso** che viene fatto tra sicurezza ed efficienza.

Client Server

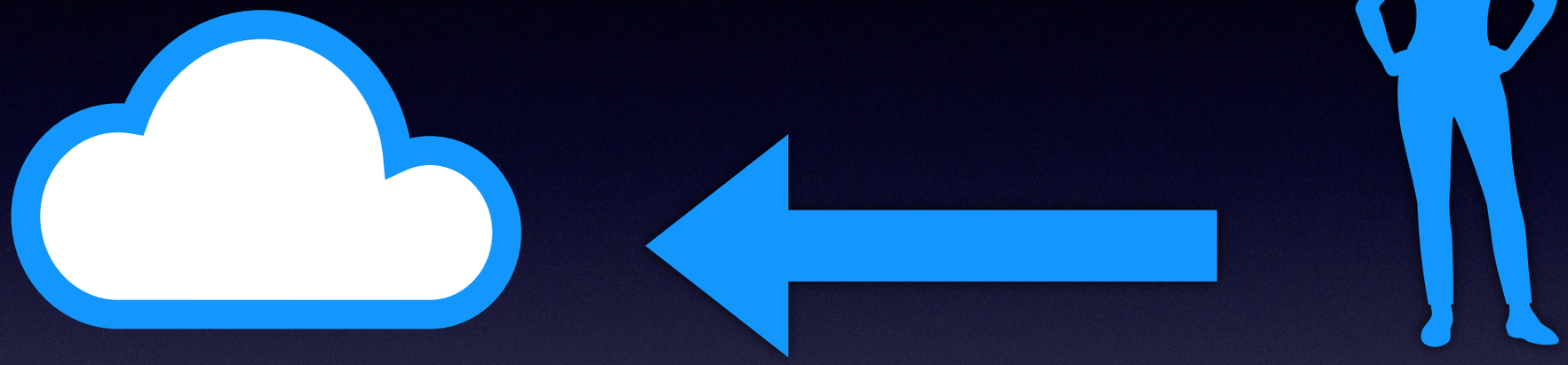
- Nell’architettura client-server i dati vengono trasmessi dal client al server attraverso internet;
- Internet rappresenta un canale di comunicazione insicuro, facilmente intercettabile da terze parti;
- Per questo è importante proteggere i dati durante questa fase di trasferimento.

End-to-End Encryption (E2EE)



cloud transport user data

End-to-End Encryption (E2EE)



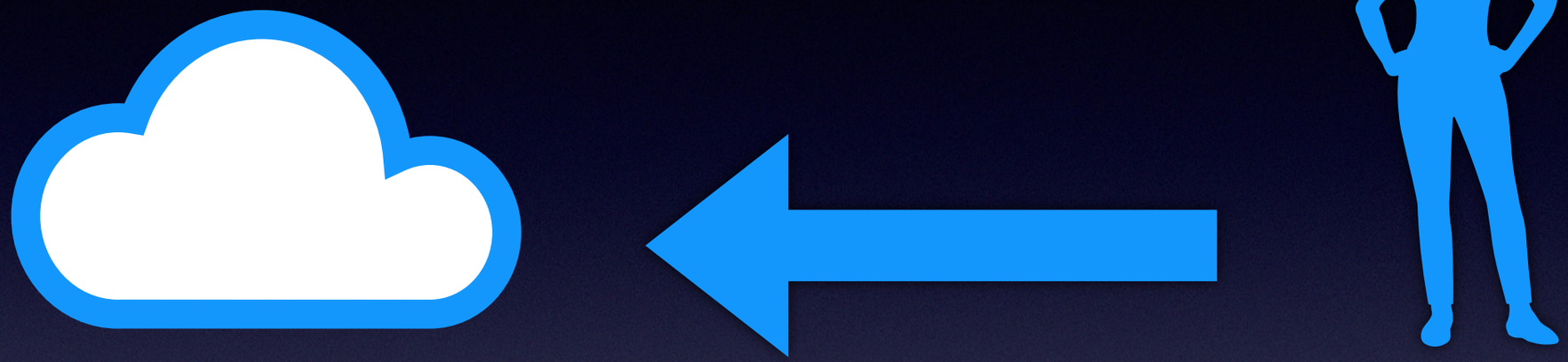
cloud

transport

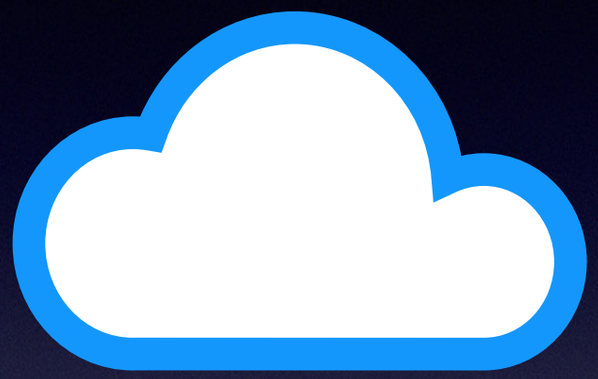
user data

data

End-to-End Encryption (E2EE)



End-to-End Encryption (E2EE)



cloud

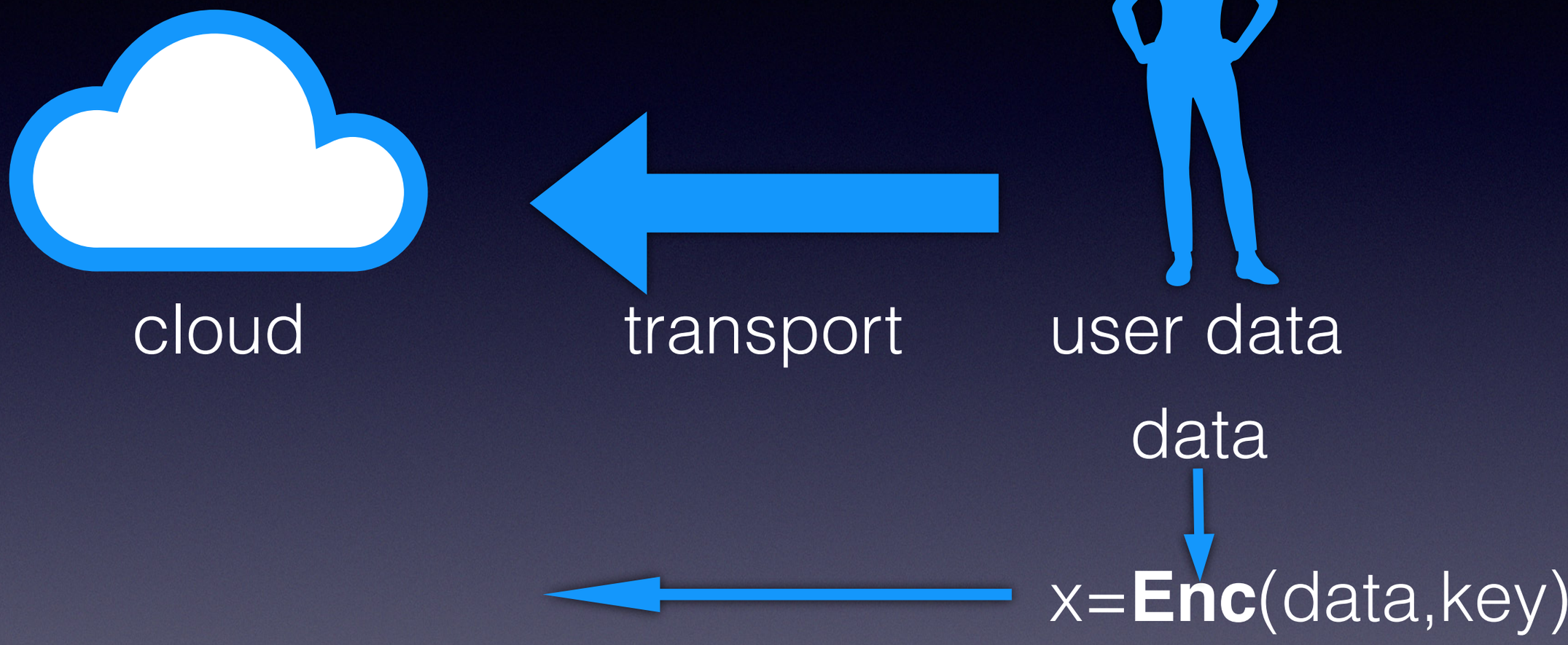
transport

user data

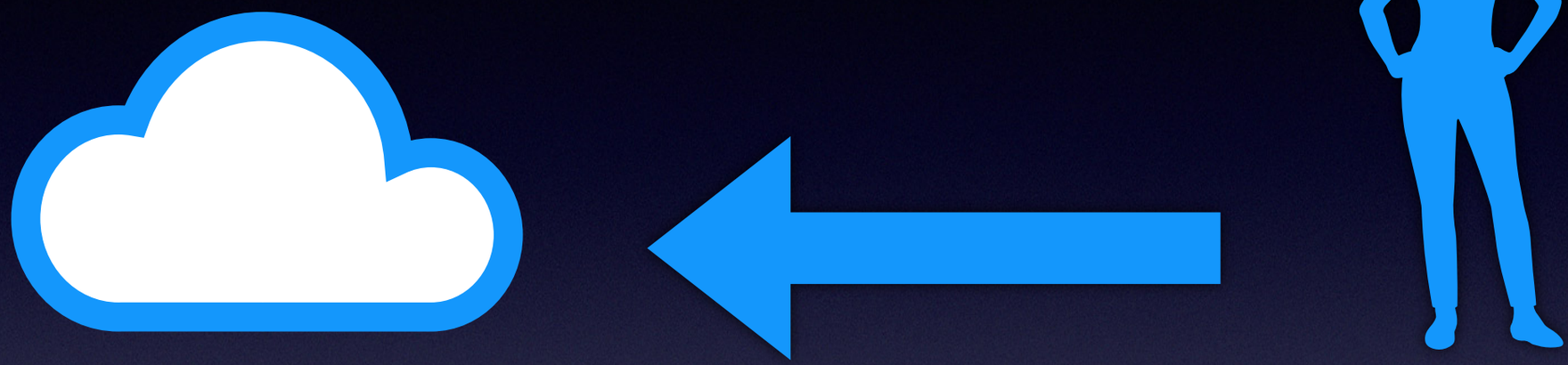
data

x=**Enc**(data,key)

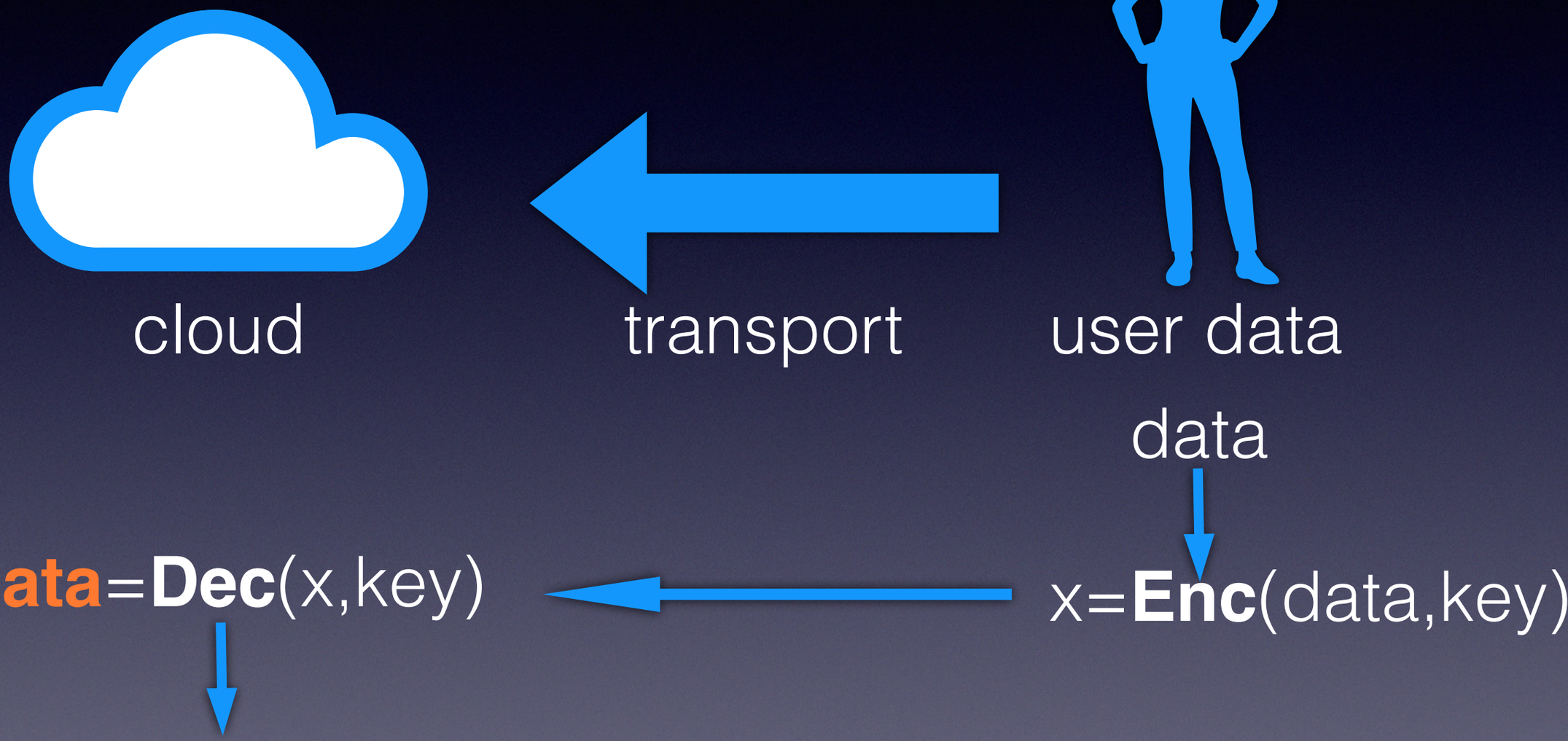
End-to-End Encryption (E2EE)



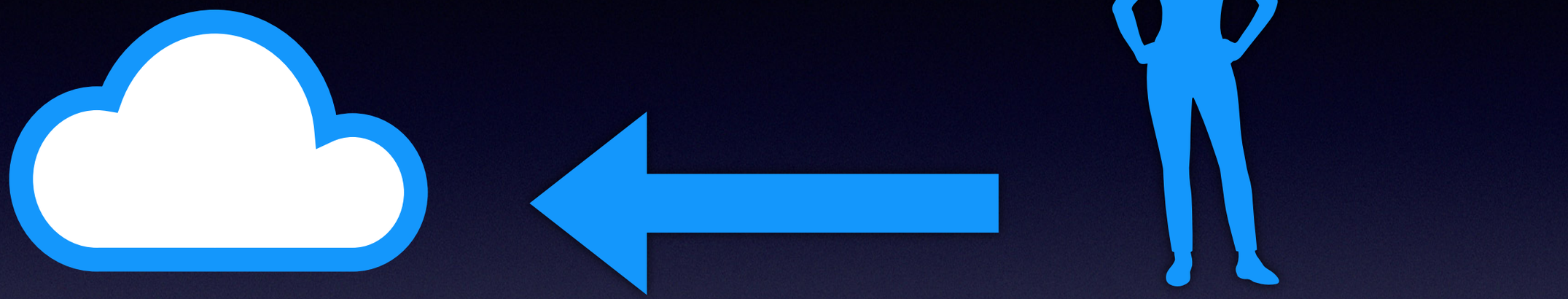
End-to-End Encryption (E2EE)



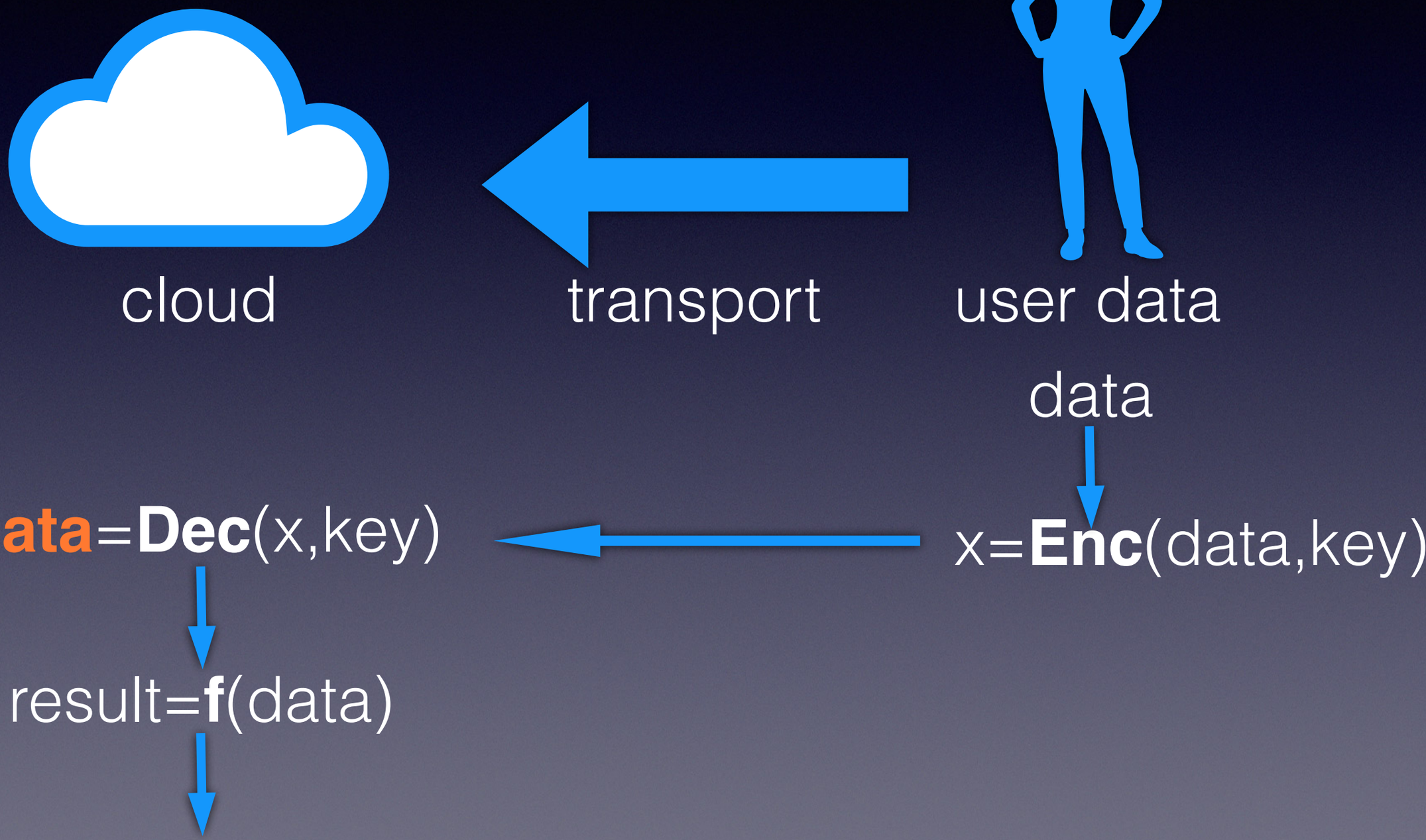
End-to-End Encryption (E2EE)



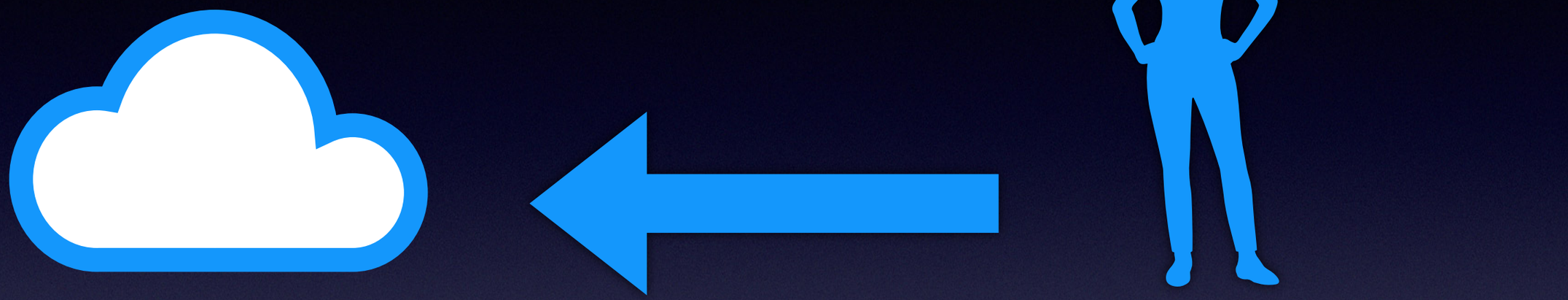
End-to-End Encryption (E2EE)



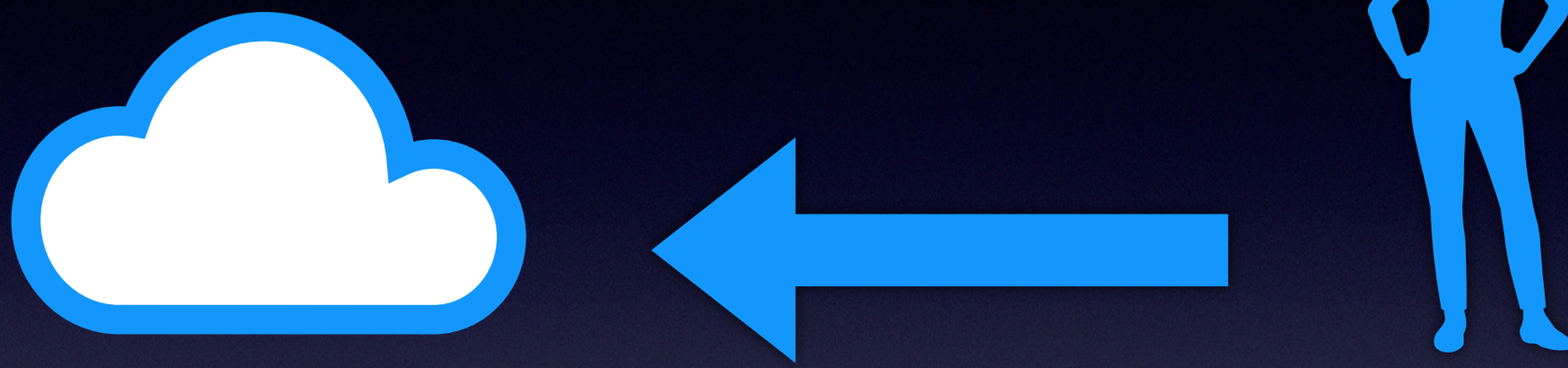
End-to-End Encryption (E2EE)



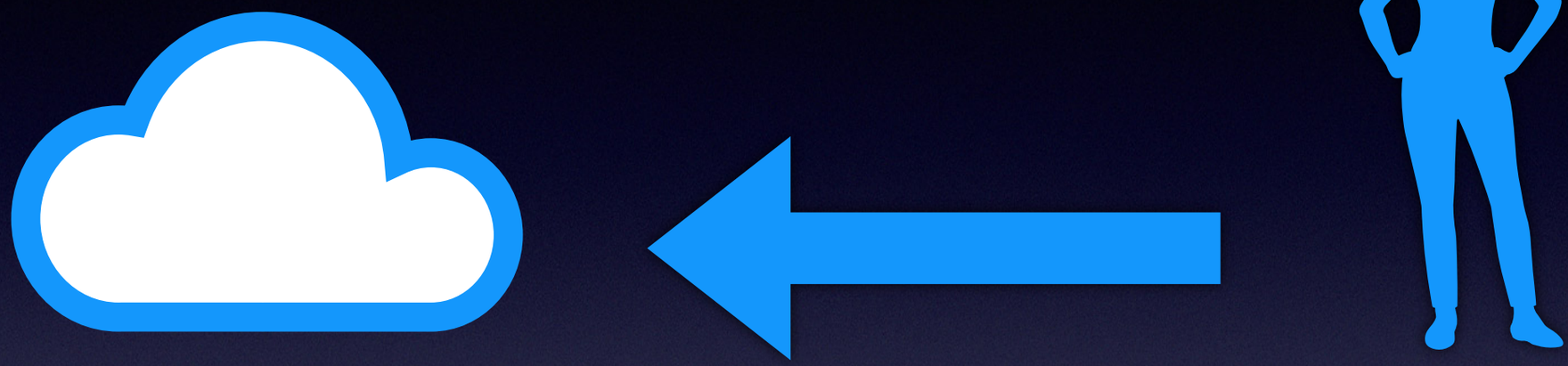
End-to-End Encryption (E2EE)



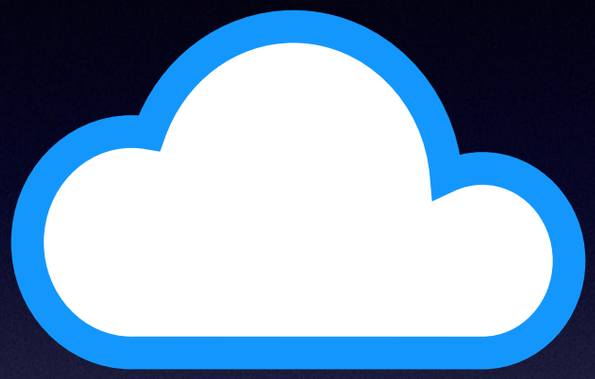
End-to-End Encryption (E2EE)



End-to-End Encryption (E2EE)

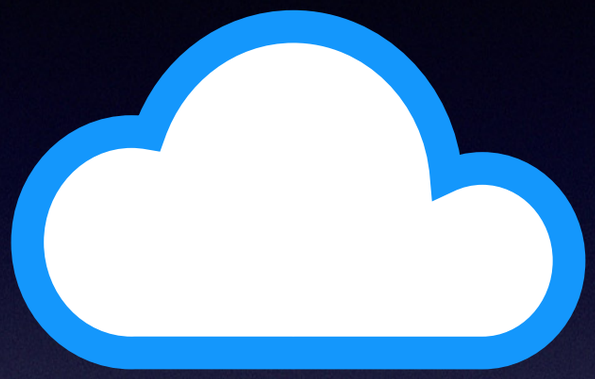


Perdere la privacy



cloud

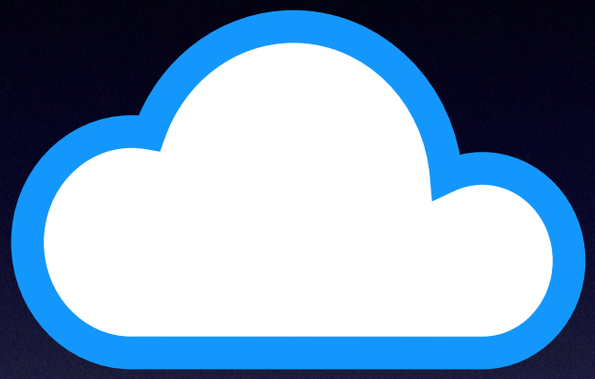
Perdere la privacy



cloud

data=**Dec**(x,key)

Perdere la privacy

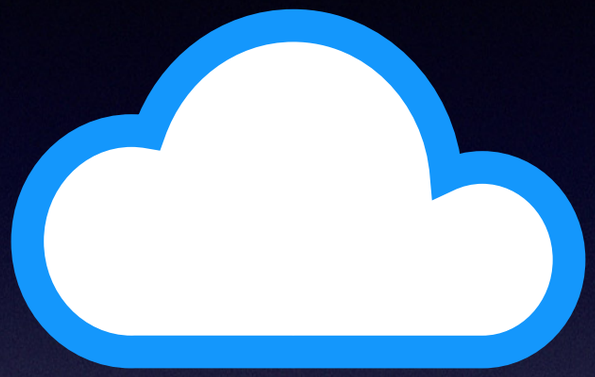


cloud

Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

data=**Dec**(x,key)

Perdere la privacy



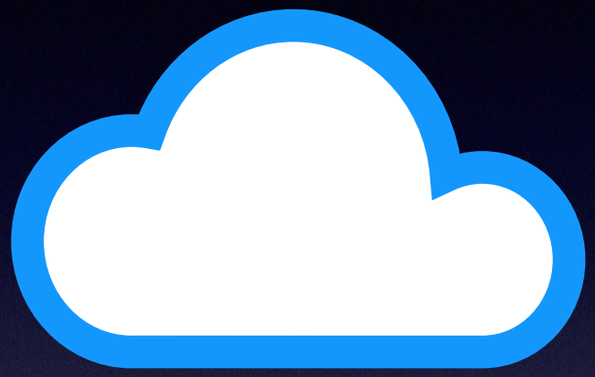
cloud

Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

- il fornitore si può avvalere a sua volta di altri fornitori;

data=**Dec**(x,key)

Perdere la privacy



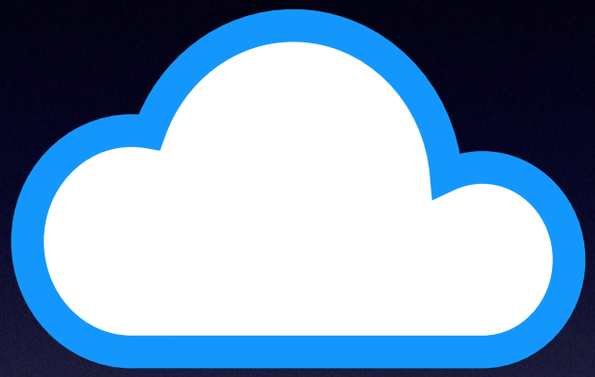
cloud

data=**Dec**(x,key)

Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

- il fornitore si può avvalere a sua volta di altri fornitori;
- può subire attacchi che sottraggono i dati;

Perdere la privacy



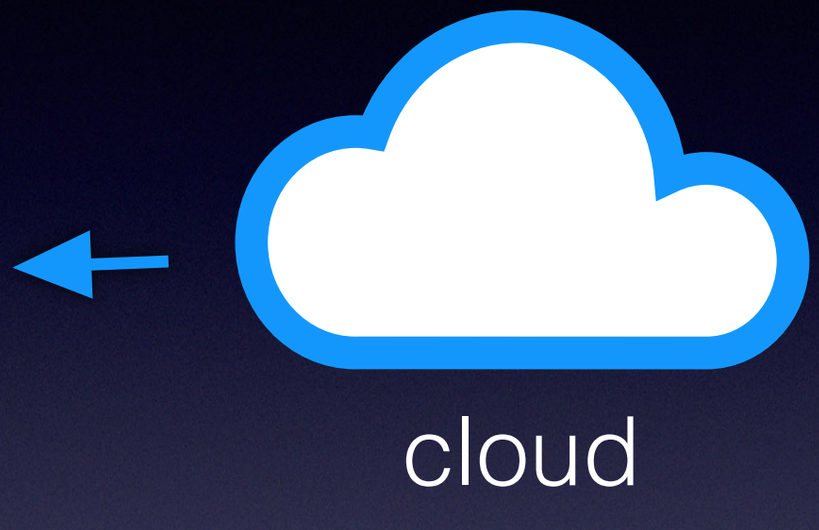
cloud

data=**Dec**(x,key)

Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

- il fornitore si può avvalere a sua volta di altri fornitori;
- può subire attacchi che sottraggono i dati;
- può ricevere pressioni dai governi.

Perdere la privacy



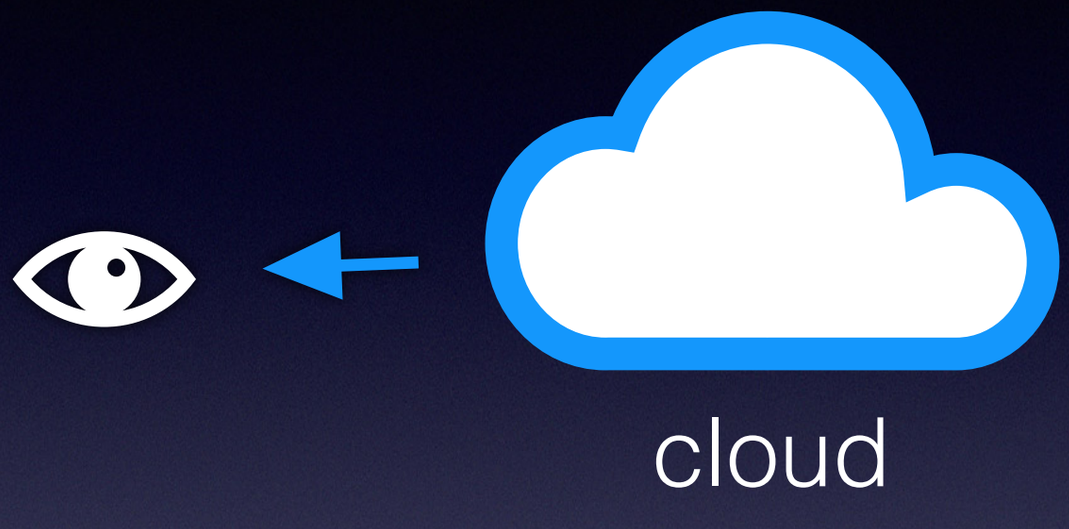
cloud

data=**Dec**(x,key)

Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

- il fornitore si può avvalere a sua volta di altri fornitori;
- può subire attacchi che sottraggono i dati;
- può ricevere pressioni dai governi.

Perdere la privacy



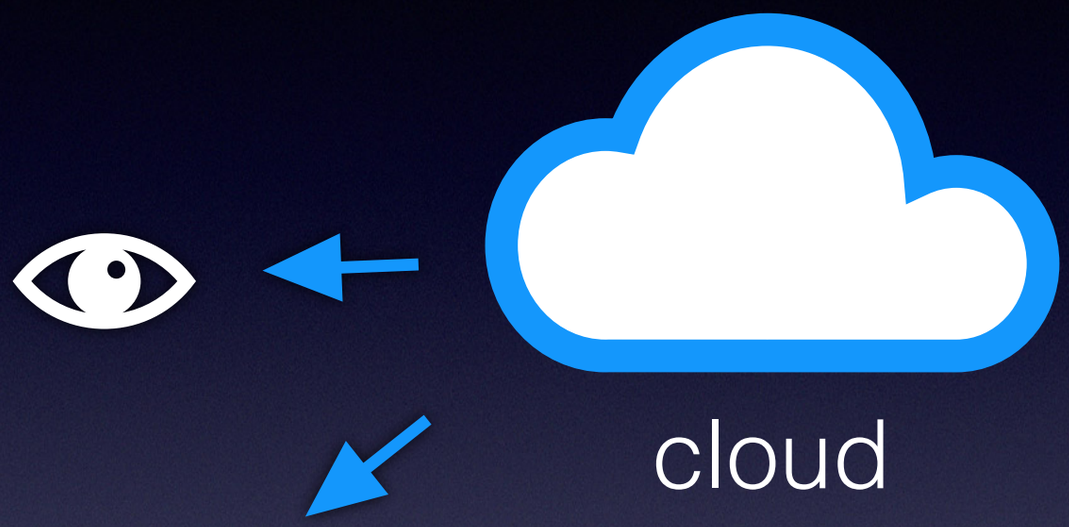
cloud

data=**Dec**(x,key)

Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

- il fornitore si può avvalere a sua volta di altri fornitori;
- può subire attacchi che sottraggono i dati;
- può ricevere pressioni dai governi.

Perdere la privacy

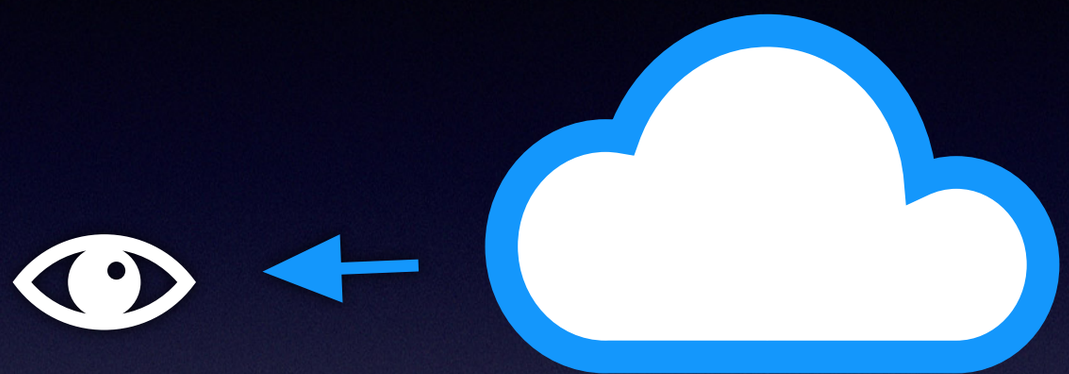


data=**Dec**(x,key)

Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

- il fornitore si può avvalere a sua volta di altri fornitori;
- può subire attacchi che sottraggono i dati;
- può ricevere pressioni dai governi.

Perdere la privacy



Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

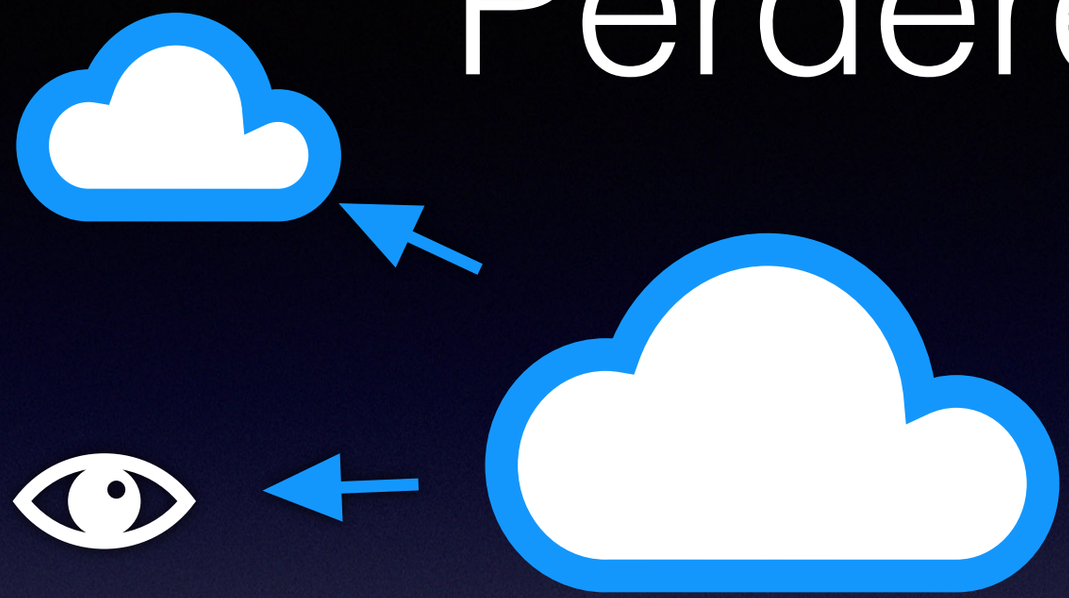
- il fornitore si può avvalere a sua volta di altri fornitori;
- può subire attacchi che sottraggono i dati;
- può ricevere pressioni dai governi.

Perdere la privacy

Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

- il fornitore si può avvalere a sua volta di altri fornitori;
- può subire attacchi che sottraggono i dati;
- può ricevere pressioni dai governi.

Perdere la privacy



Quando il dato è a disposizione del fornitore di servizi, l'utente ha già perso il controllo:

- il fornitore si può avvalere a sua volta di altri fornitori;
- può subire attacchi che sottraggono i dati;
- può ricevere pressioni dai governi.

Elaborazioni

- Il fornitore mette in campo tutti i mezzi per preservare la privacy, proteggendo i dati dell'utente
- Il punto essenziale è che per poter fornire il servizio deve accedere ai dati in chiaro
- Dati molto sensibili: ad esempio i **dispositivi indossabili** che effettuano il monitoraggio dei parametri biologici, le telecamere di **sorveglianza** dei sistemi di allarme di casa…
- Vorremmo che i provider fornissero **elaborazioni** su questi dati (tipicamente per mezzo di algoritmi di *machine learning*) senza che avessero accesso agli stessi.

Proteggere la privacy

- E’ un *paradosso* voler mantenere la **privacy** e il controllo dei propri dati ed utilizzare fornitori di **servizi** su questi dati ?

Non Decifrare Mai !!

Non Decifrare Mai !!

- il lato **Client cifra** i dati da conferire al Server;

Non Decifrare Mai !!

- il lato **Client cifra** i dati da conferire al Server;
- la **chiave** resta segreta, conosciuta solo all’utente;

Non Decifrare Mai !!

- il lato **Client cifra** i dati da conferire al Server;
- la **chiave** resta segreta, conosciuta solo all'utente;
- i **dati** arrivano al fornitore di servizi **cifrati** e a differenza dell'E2EE, dal lato server, la chiave non è nota;

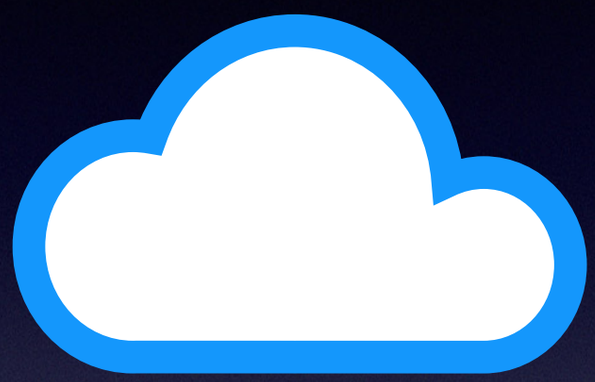
Non Decifrare Mai !!

- il lato **Client cifra** i dati da conferire al Server;
- la **chiave** resta segreta, conosciuta solo all'utente;
- i **dati** arrivano al fornitore di servizi **cifrati** e a differenza dell'E2EE, dal lato server, la chiave non è nota;
- lato server restano solo dati cifrati.

Non Decifrare Mai !!

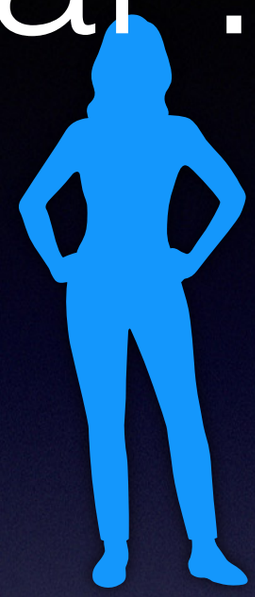
- il lato **Client cifra** i dati da conferire al Server;
- la **chiave** resta segreta, conosciuta solo all'utente;
- i **dati** arrivano al fornitore di servizi **cifrati** e a differenza dell'E2EE, dal lato server, la chiave non è nota;
- lato server restano solo dati cifrati.
- In questo approccio possiamo pensare di cifrare i dati e poi dimenticare/**distruggere la chiave**…

Non Decifrare Mai !!



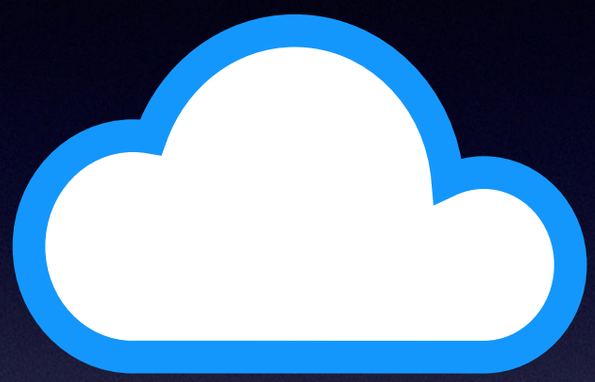
cloud

transport



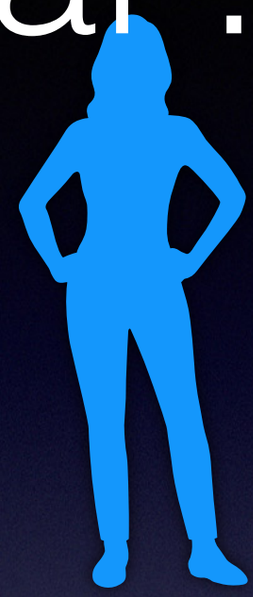
user data

Non Decifrare Mai !!



cloud

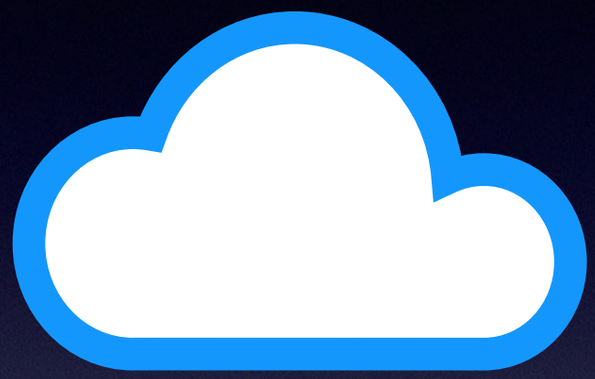
transport



user data

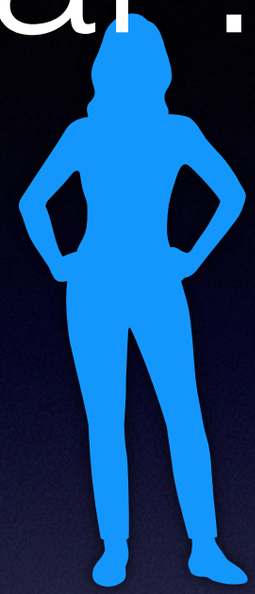
data

Non Decifrare Mai !!



cloud

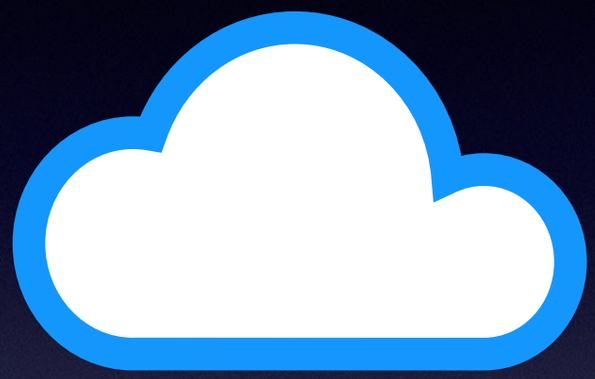
transport



user data

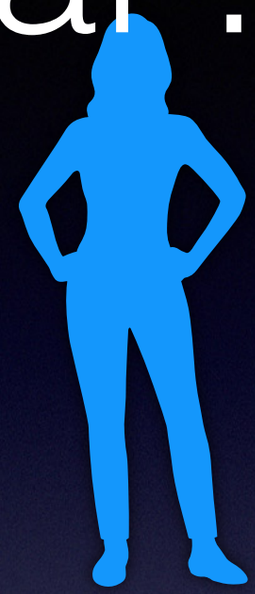
data

Non Decifrare Mai !!



cloud

transport

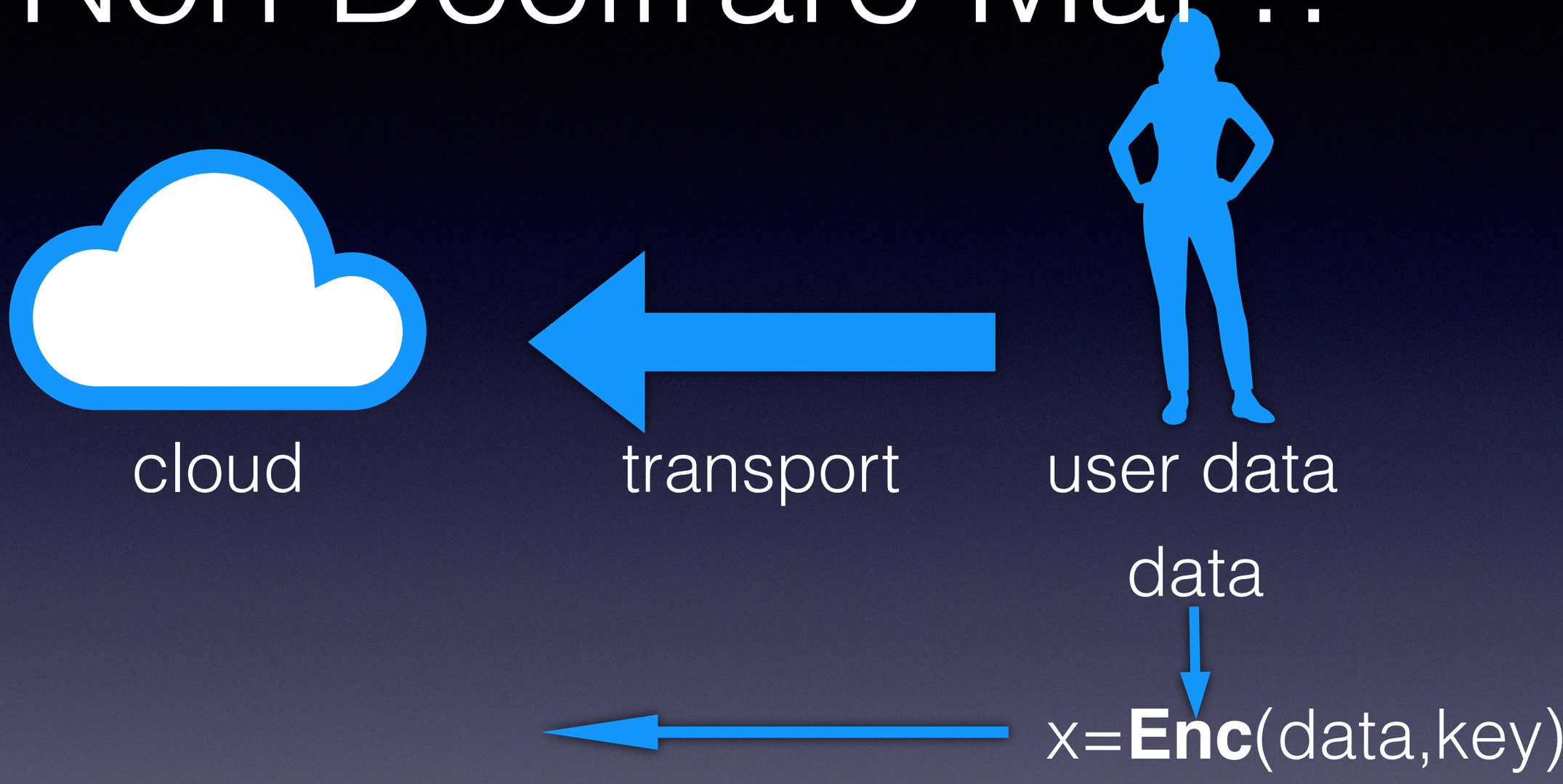


user data

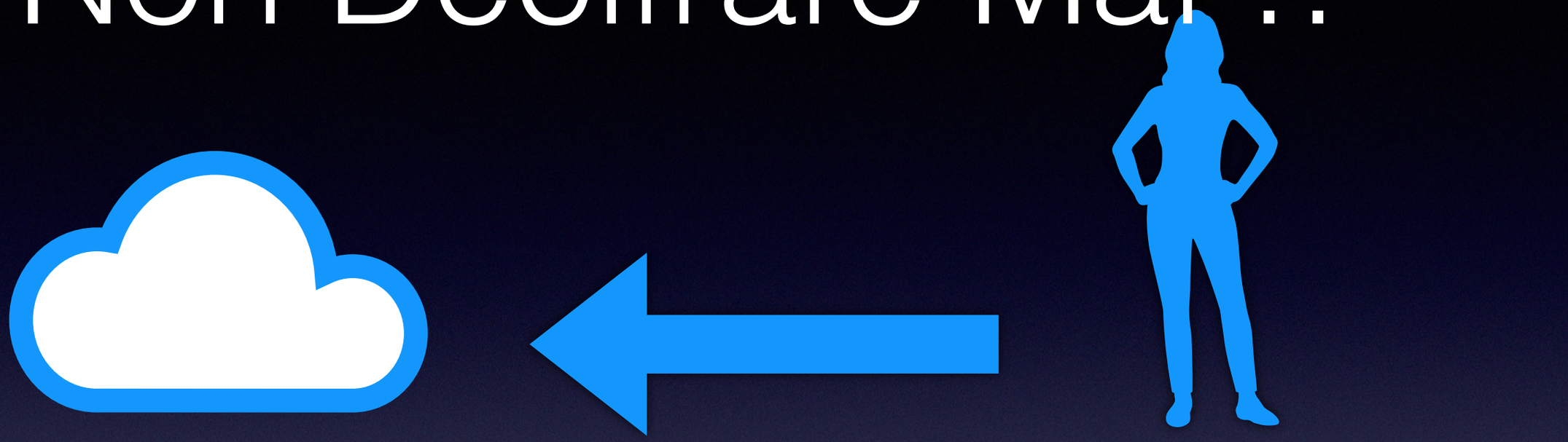
data

x=**Enc**(data,key)

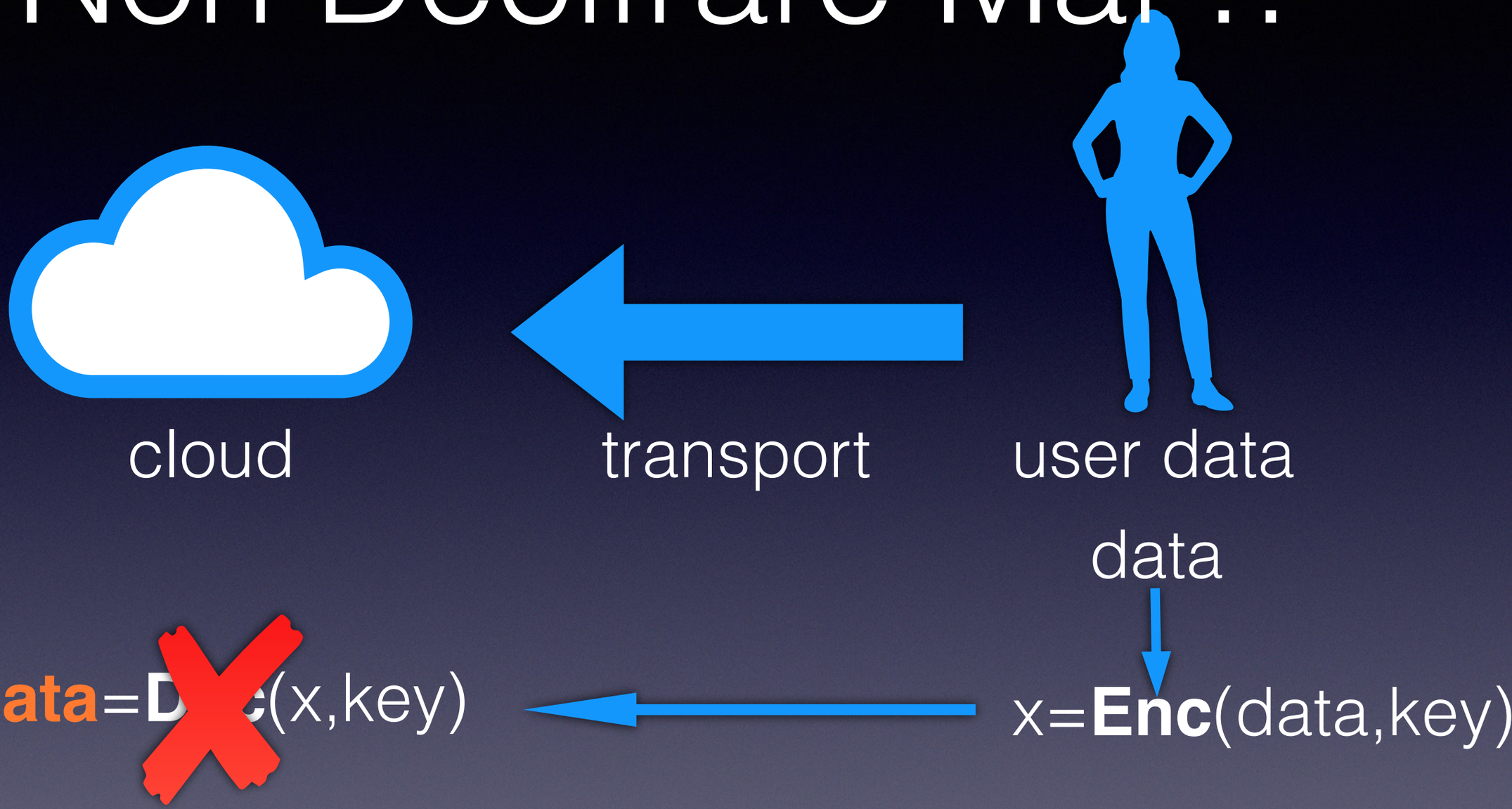
Non Decifrare Mai !!



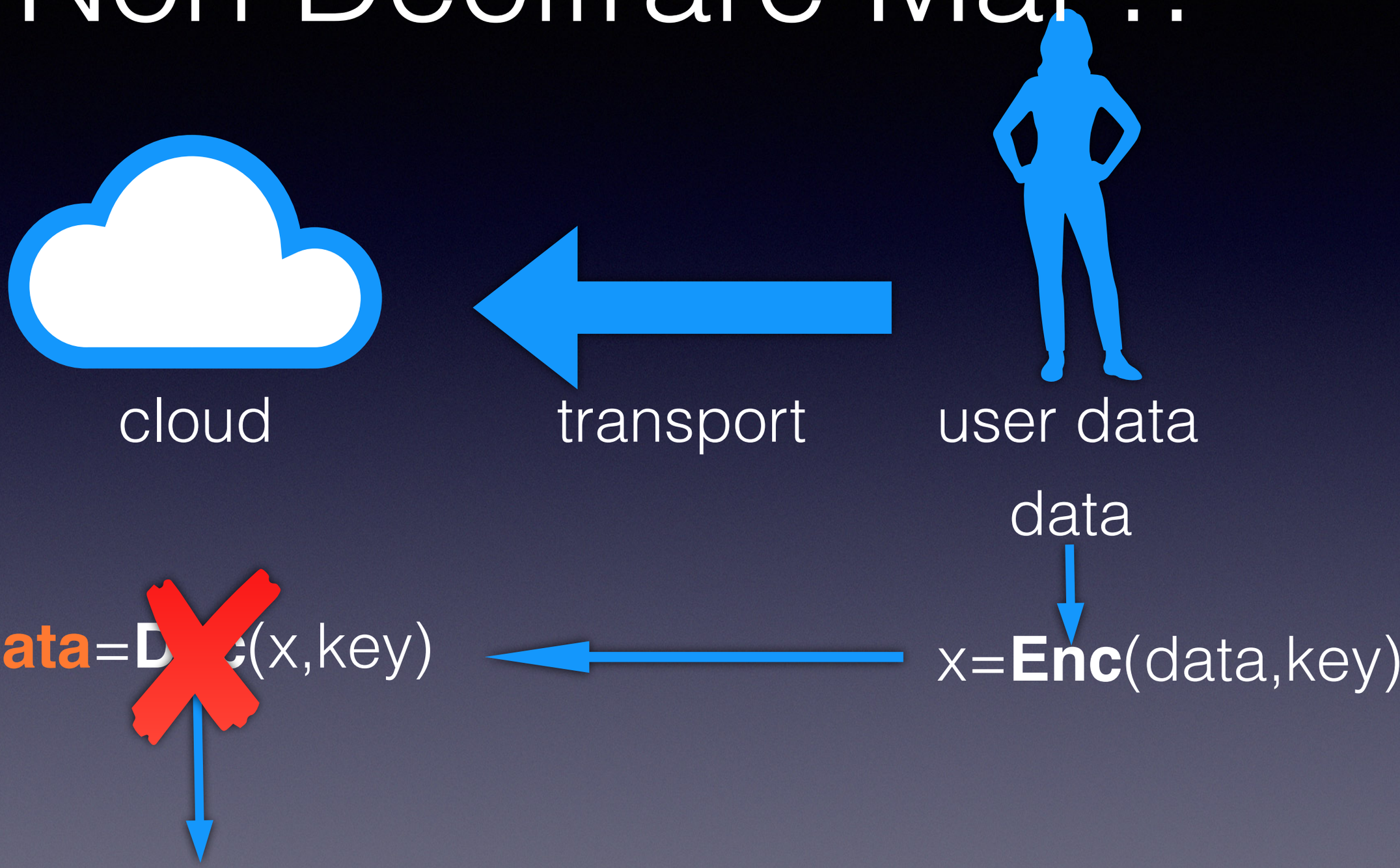
Non Decifrare Mai !!



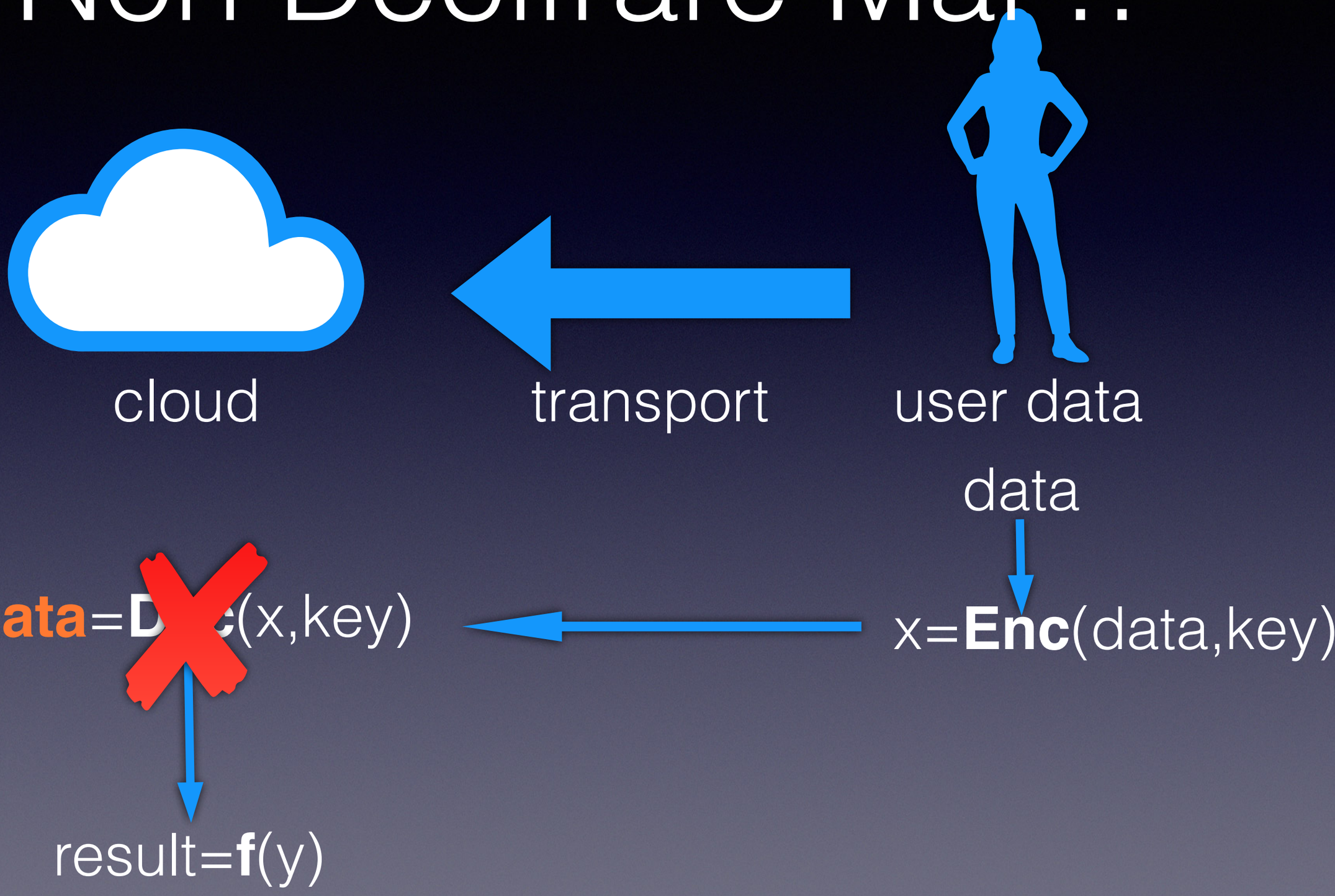
Non Decifrare Mai !!



Non Decifrare Mai !!



Non Decifrare Mai !!



Nuovo Scenario

- Conferire solo dati cifrati al cloud…
- …il provider di servizi elabora utilizzando i dati cifrati senza mai decifrare.
- Cosa è f(**Enc**(data,key)) ?

Homomorphic Encryption

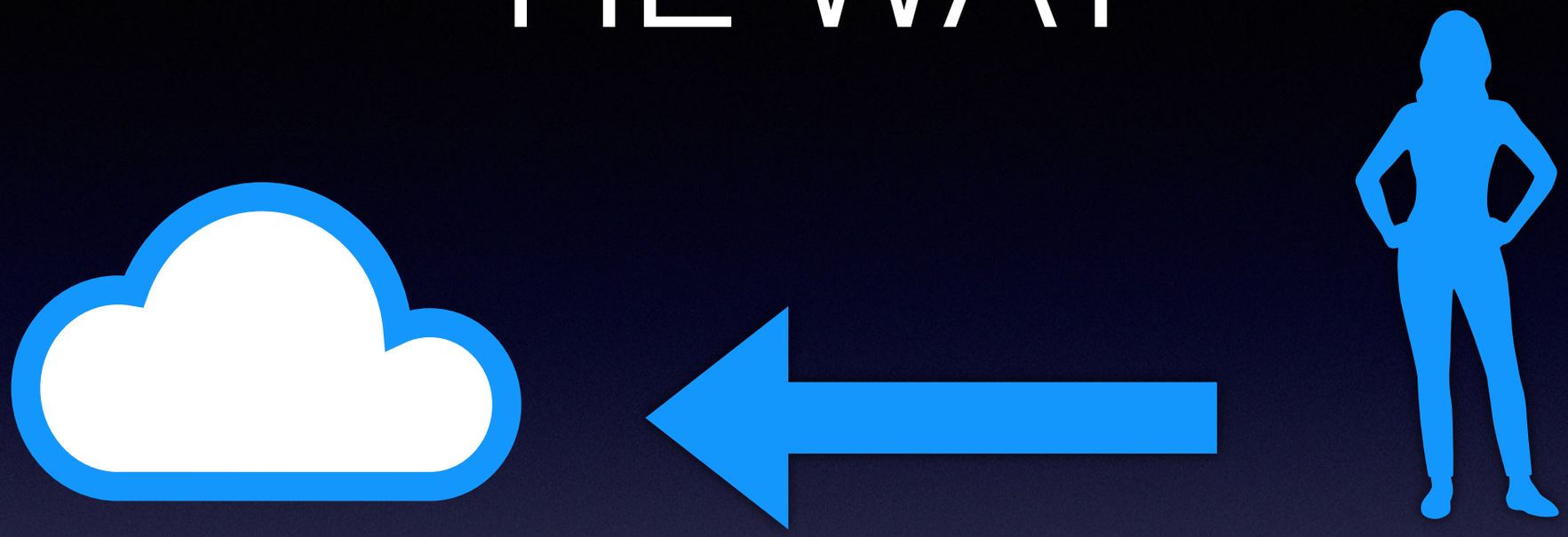
- Per casi particolari di funzioni di cifratura **Enc**() e di funzioni f() particolari vale la proprietà:

Enc(f(data),key) = f(**Enc**(data,key))

- dunque il server può effettuare l'elaborazione sul dato cifrato e rinviare il risultato cifrato al client che eventualmente può decrittare il risultato.

Dec(**Enc**(f(data),key),key)=f(data).

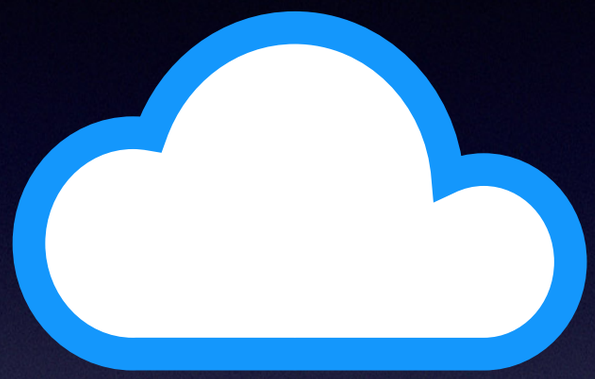
HE WAY



cloud transport user data

f(x)=f(**Enc**(data,key))=**Enc**(f(data),key)
Dec(f(x),key)=**Dec**(**Enc**(f(data),key))=f(data)

HE WAY



cloud

transport

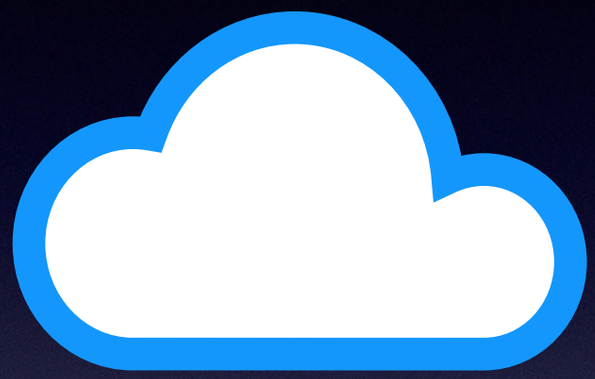
user data

data

f(x)=f(**Enc**(data,key))=**Enc**(f(data),key)

Dec(f(x),key)=**Dec**(**Enc**(f(data),key))=f(data)

HE WAY



cloud

transport

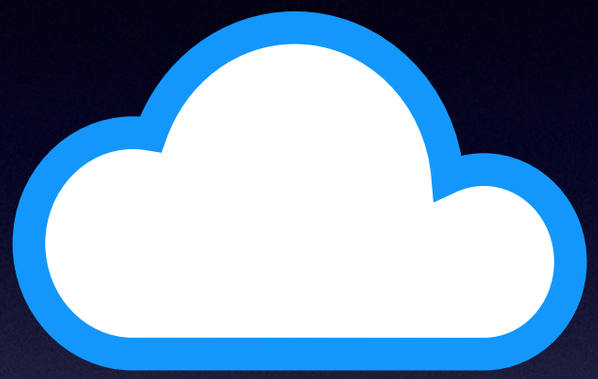
user data

data

f(x)=f(**Enc**(data,key))=**Enc**(f(data),key)

Dec(f(x),key)=**Dec**(**Enc**(f(data),key))=f(data)

HE WAY



cloud

transport

user data

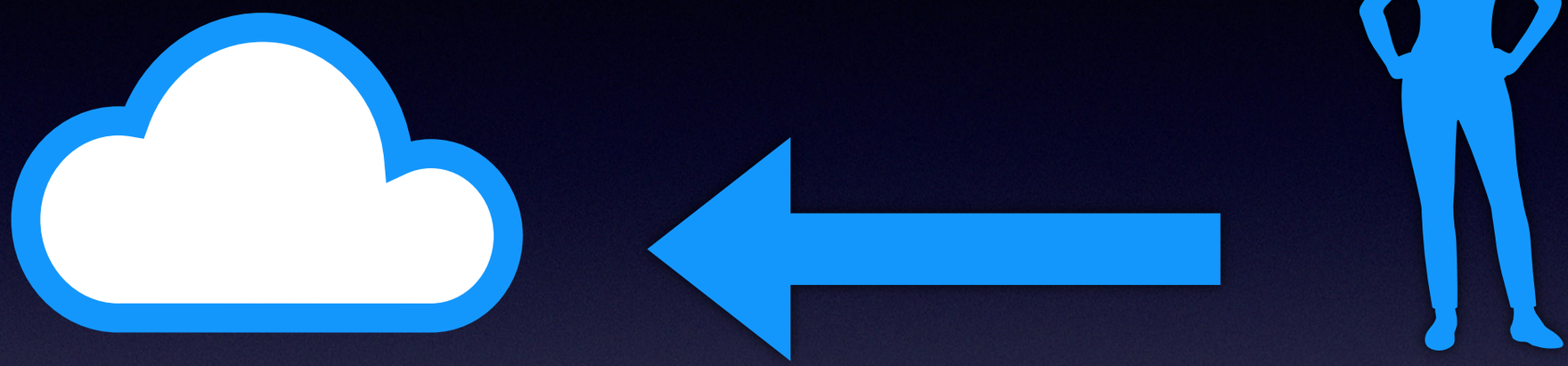
data

x=**Enc**(data,key)

f(x)=f(**Enc**(data,key))=**Enc**(f(data),key)

Dec(f(x),key)=**Dec**(**Enc**(f(data),key))=f(data)

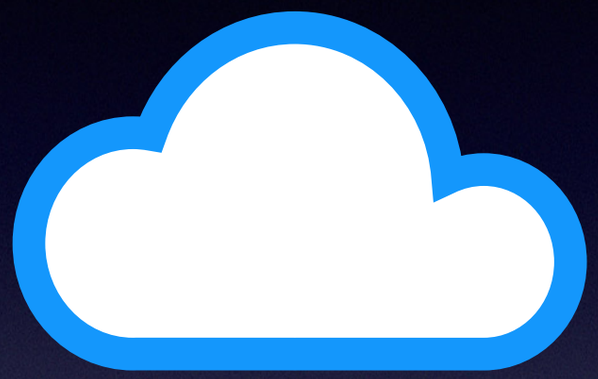
HE WAY



f(x)=f(**Enc**(data,key))=**Enc**(f(data),key)

Dec(f(x),key)=**Dec**(**Enc**(f(data),key))=f(data)

HE WAY



cloud

transport

user data

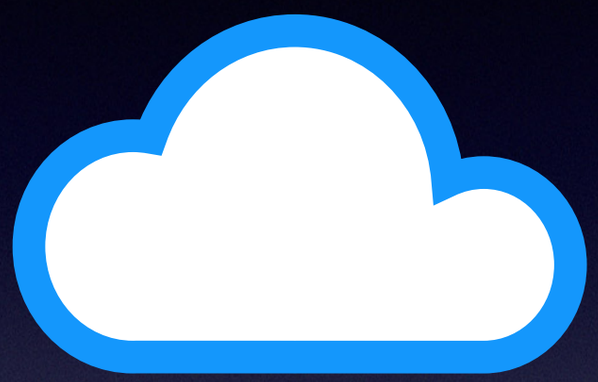
data

x=**Enc**(data,key)

f(x)=f(**Enc**(data,key))=**Enc**(f(data),key)

Dec(f(x),key)=**Dec**(**Enc**(f(data),key))=f(data)

HE WAY



cloud transport user data

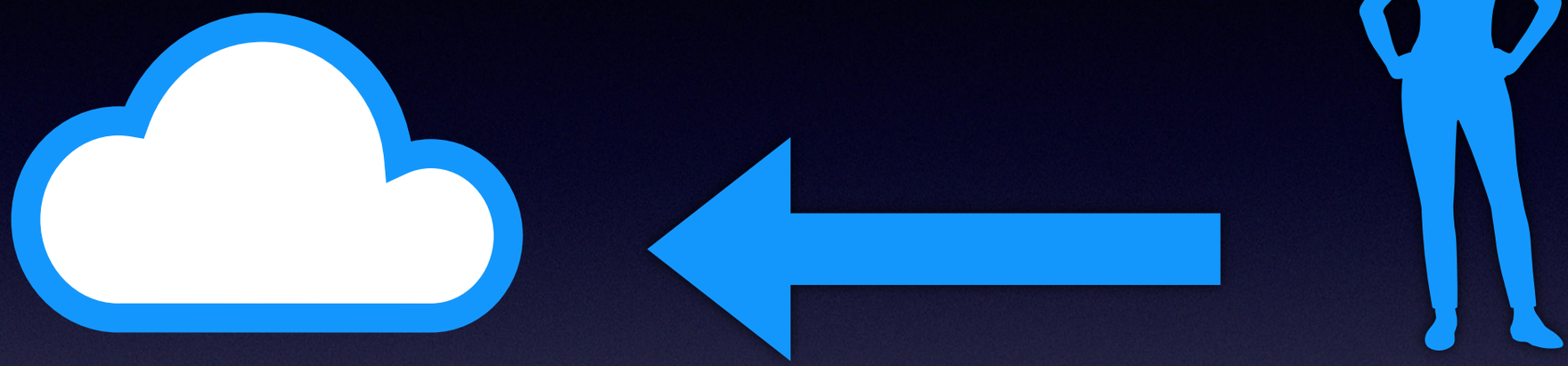
data

Enc(result,key)=f(x) x=**Enc**(data,key)

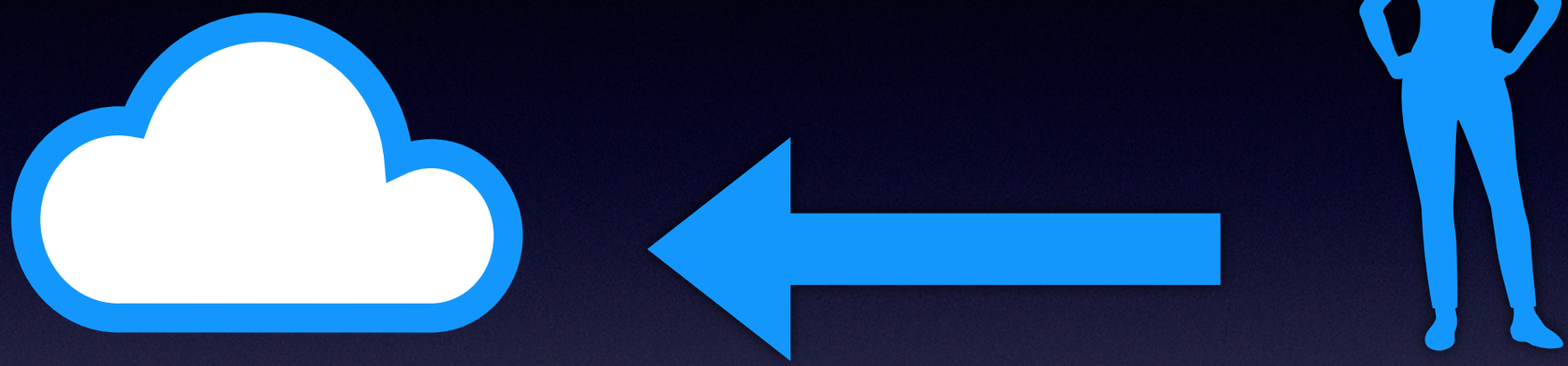
f(x)=f(**Enc**(data,key))=**Enc**(f(data),key)

Dec(f(x),key)=**Dec**(**Enc**(f(data),key))=f(data)

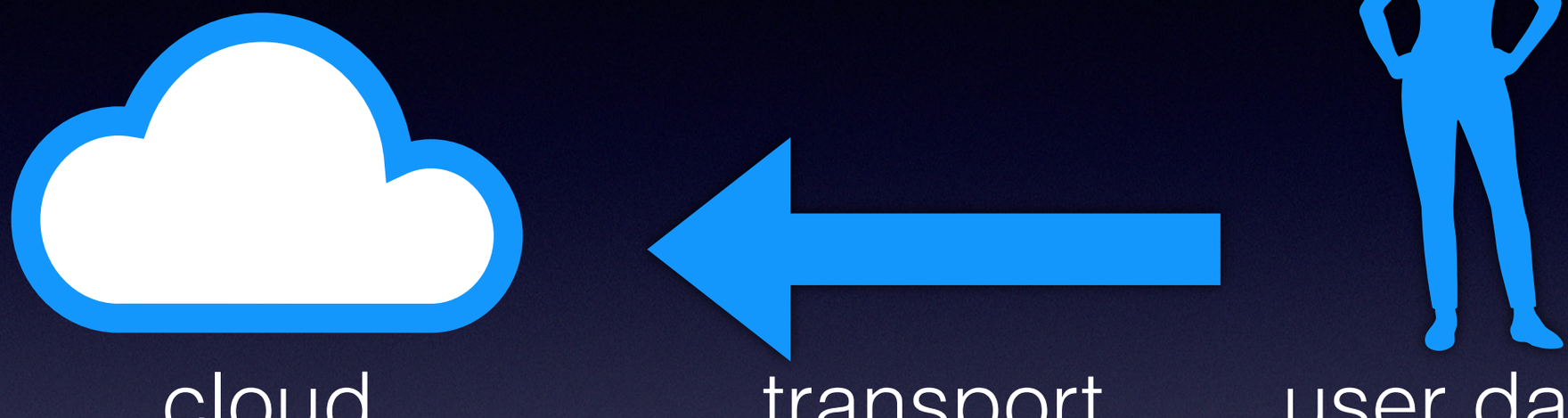
HE WAY


cloud
transport
user data
data
Enc(result,key)=f(x)
x=Enc(data,key)
f(x)=f(Enc(data,key))=Enc(f(data),key)
Dec(f(x),key)=Dec(Enc(f(data),key))=f(data)
y=Enc(result,key)

HE WAY


cloud
transport
user data
data
Enc(result,key)=f(x)
x=Enc(data,key)
f(x)=f(Enc(data,key))=Enc(f(data),key)
Dec(f(x),key)=Dec(Enc(f(data),key))=f(data)
y=Enc(result,key)

HE WAY



data

Enc(result,key)=f(x)

x=**Enc**(data,key)

f(x)=f(**Enc**(data,key))=**Enc**(f(data),key)

Dec(f(x),key)=**Dec**(**Enc**(f(data),key))=f(data)

y=**Enc**(result,key)

result=**Dec**(y,key)

Esistono sistemi per HE

- Nel 1978 Rivest aveva posto la questione se fosse possibile trovare per qualsiasi funzione un crittosistema con la proprietà della HE:

Enc(f(data),key) = f(**Enc**(data,key))

- Negli anni sono state trovate funzioni f con la proprietà HE rispetto a particolari crittosistemi: addizione modulo 2, addizione con interi piccoli e moltiplicazione modulo p (ElGamal1984), addizione di interi grandi (Paillier1999), successore e moltiplicazione (Brakereski et al. 2005).

Full Homomorphic Encryption

- La questione del 1978 viene detta Full Homomorphic Encryption (FHE)
- La prova che esistono crittosistemi che sono HE per ogni funzione f() arriva nel 2012 da Craig Gentry.
- Si basa sulla possibilità di avere un crittosistema di HE per somma e prodotto, da cui deriva la possibilità di avere HE per i polinomi e via approssimazione per qualsiasi f().