



UNIVERSITÀ
DEGLI STUDI
FIRENZE

ANNO ACCADEMICO 2016/2017

Implementazione e testing di un cluster di database su una rete di pari

**Scuola di Scienze Matematiche,
Fisiche e Naturali**

Corso di Laurea in Informatica

Relatore
Prof. Lorenzo Bettini

Candidato
Linda Luciano

Scopo del progetto

La replica di dati è un trasferimento dati unidirezionale da uno o più nodi che permette:

- aumento di affidabilità del sistema, in modo da renderlo tollerante ai guasti
- miglioramento delle prestazioni del sistema, aumentandone la scalabilità

Scopo del progetto

Trovare compromesso tra:

- Numero di repliche
- Sicurezza del dato
- Buone performance

Qual è il modo più proficuo per eseguire la replica di un file
in upload?





Tecnologie



➤ Fault tolerance (tolleranza ai guasti)

in caso di guasto del singolo server, il cluster offre un'alternativa a cui connettersi

➤ Load balancing (bilanciamento del carico)

consente agli utenti di essere assegnati automaticamente al server con il minor carico

Principali vantaggi di una rete di pari (*peer-to-peer network*):

- Se un dispositivo collegato interrompe la connessione, il servizio non termina
- Non si basa esclusivamente su server centrali, è più resistente in caso di guasti o colli di bottiglia del traffico
- Con l'aumentare del numero di dispositivi, aumenta la potenza della rete, poiché ogni computer aggiuntivo è disponibile per l'elaborazione dei dati



Architettura del progetto



Suddivisione di un file

Quando un file è scritto sul disco sono scritte due informazioni:

- File intero diviso in chunk
- Una parte di metadato, che rappresenta l'informazione anagrafica del dato

Suddivisione di un file

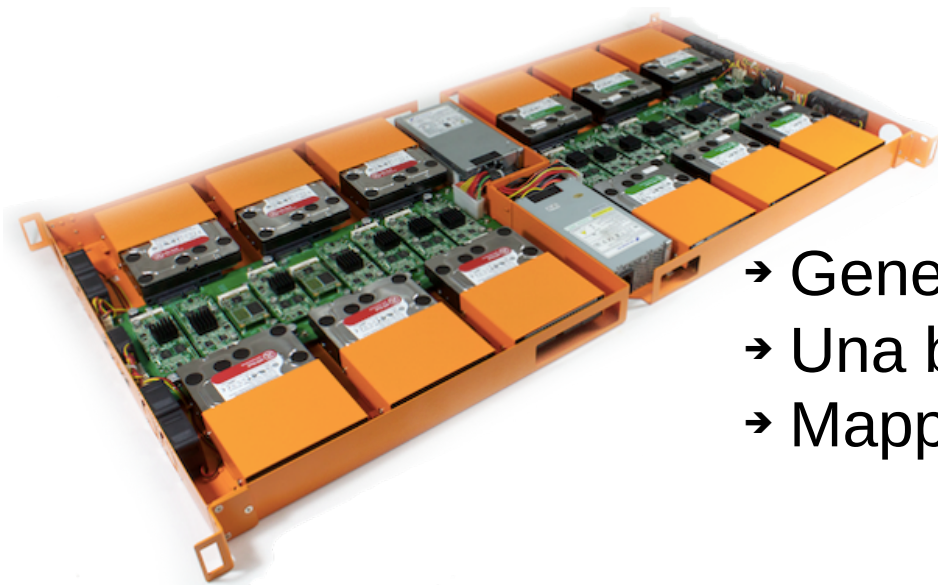
In che modo possiamo ottenere affidabilità del dato/metadato senza compromettere le performance?



Replicare i dati/metadati su almeno due board differenti in modo tale che, in casi di fault, guasti o perdite, sia possibile ottenere nuovamente il dato originale.

Replica metadato/dato

Scrittura di un file



- Generazione ID
- Una board gestisce un range di ID
- Mappa restituisce IP

Problema:
se board 01 è offline?

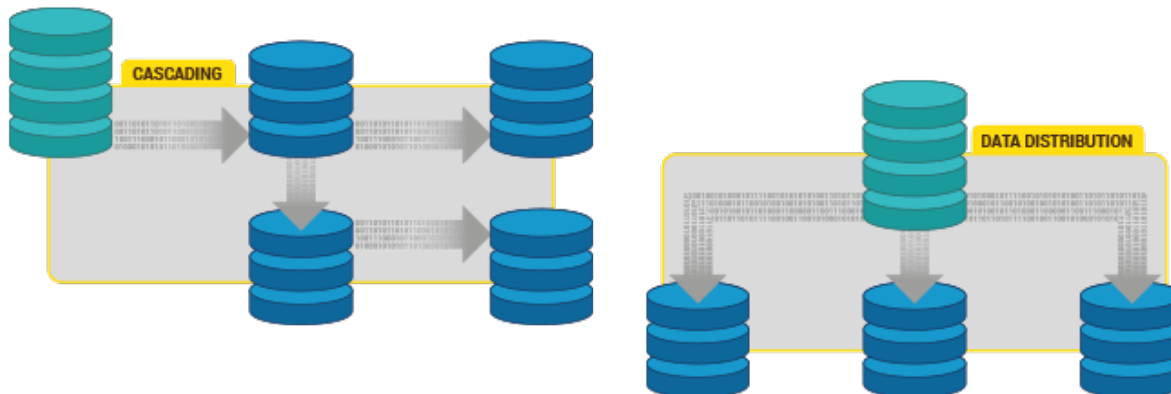
BOARD	range ID
B 01 01	0-10
B 01 02	11-20
B 01 03	21-30
B 01 01	31-40
B 01 02	41-50
B 01 03	51-60

➤ Write-Ahead Log (WAL)

- Garantisce integrità dei metadati
- In caso di crash è possibile recuperare il database usando il log

➤ Pglogical

- Replica completa o selettiva di insieme di tabelle di database ad altri nodi di un cluster



Le mappe

➤ Per i metadati:

la mappa gestisce un solo nodo per board, un solo database, quindi una sola replica

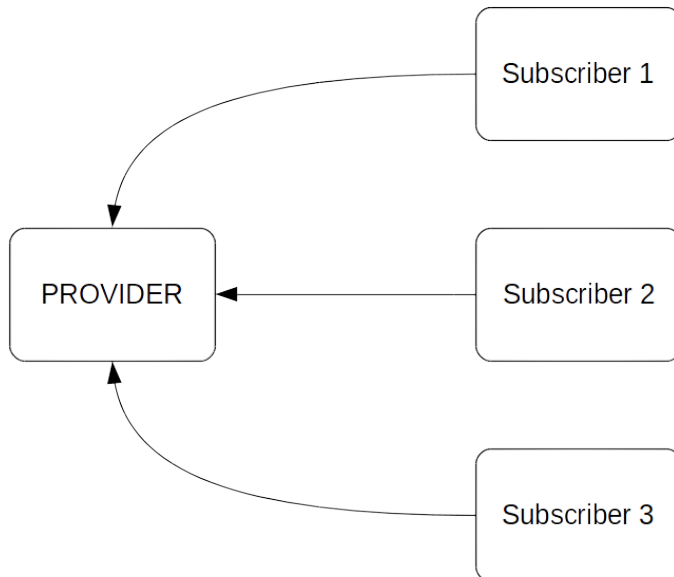
➤ Per i dati:

la mappa gestisce più nodi per board

Ogni board ha otto database di dati e uno solo di metadati.

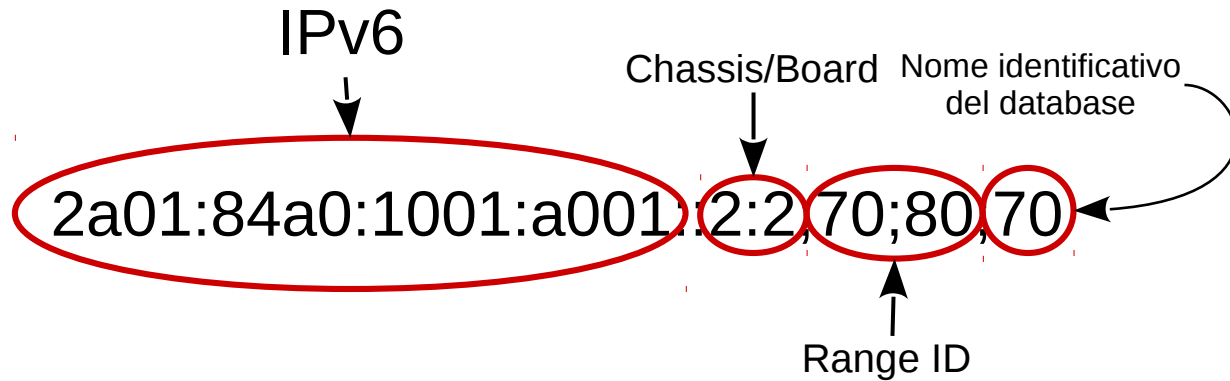
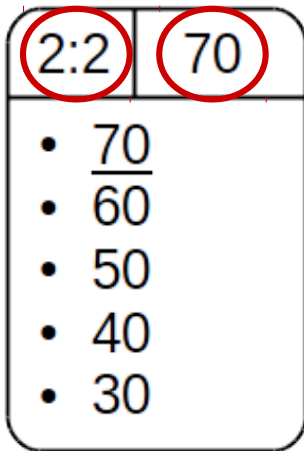


Publish - Subscribe



Utilizzando la replica in cascata, ogni nodo *Subscriber* è in contemporanea *Provider* di altri nodi *Subscriber*

Rappresentazione del nodo



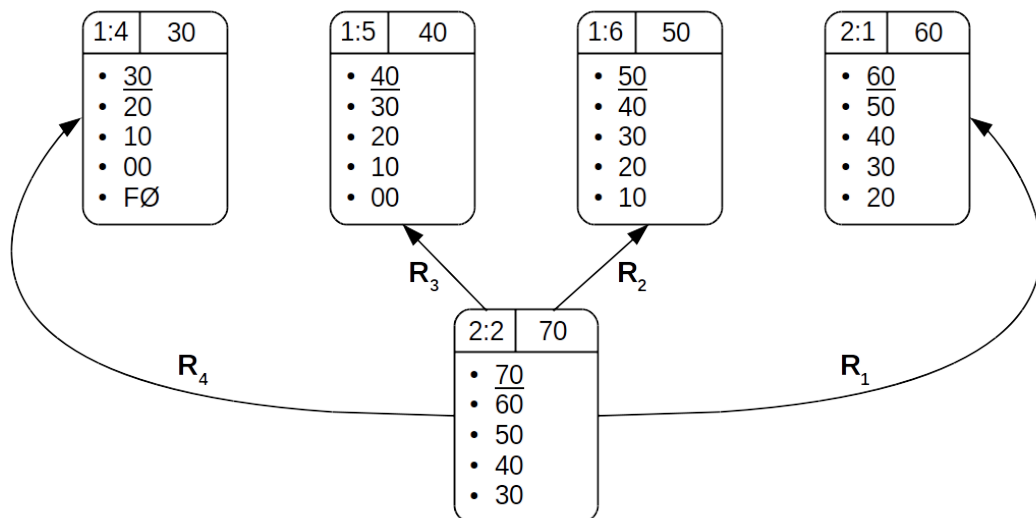
Replica asincrona

Non aspetta che lo storage primario confermi la completa scrittura del dato sul disco. Il client ha la conferma di avvenuto successo quando è stato scritto sul WAL.

Replica sincrona

Il client ha la conferma di avvenuto successo solo quando è stato effettivamente replicato il dato. Scrive la copia primaria e la replica contemporaneamente, in modo da renderli sempre sincronizzati.

Replica metadato – Set di replica



2a01:84a0:1001:a001::1:4;30;40;30

2a01:84a0:1001:a001::1:5;40;50;40

2a01:84a0:1001:a001::1:6;50;60;50

2a01:84a0:1001:a001::2:1;60;70;60

2a01:84a0:1001:a001::2:2;70;80;70

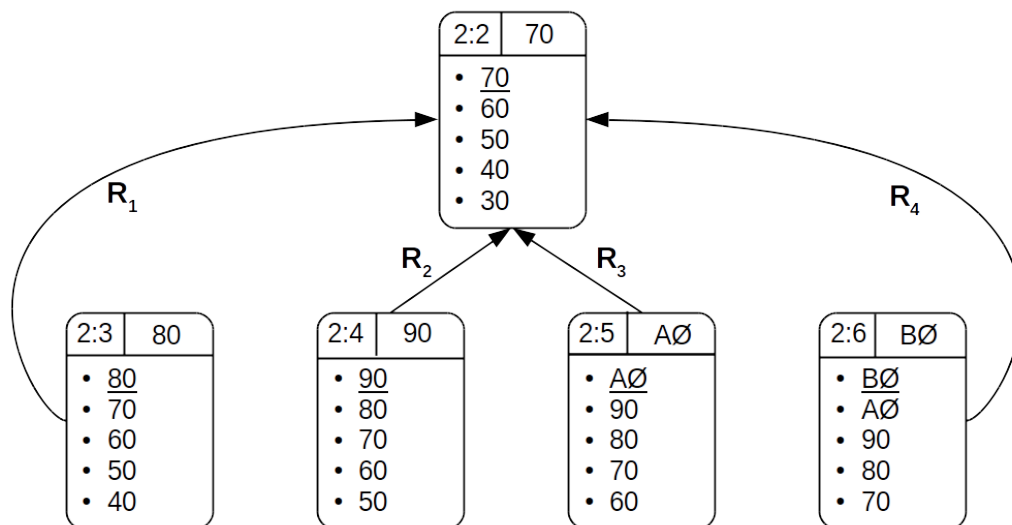
2a01:84a0:1001:a001::2:2;70;80;70

2a01:84a0:1001:a001::2:3;80;90;80

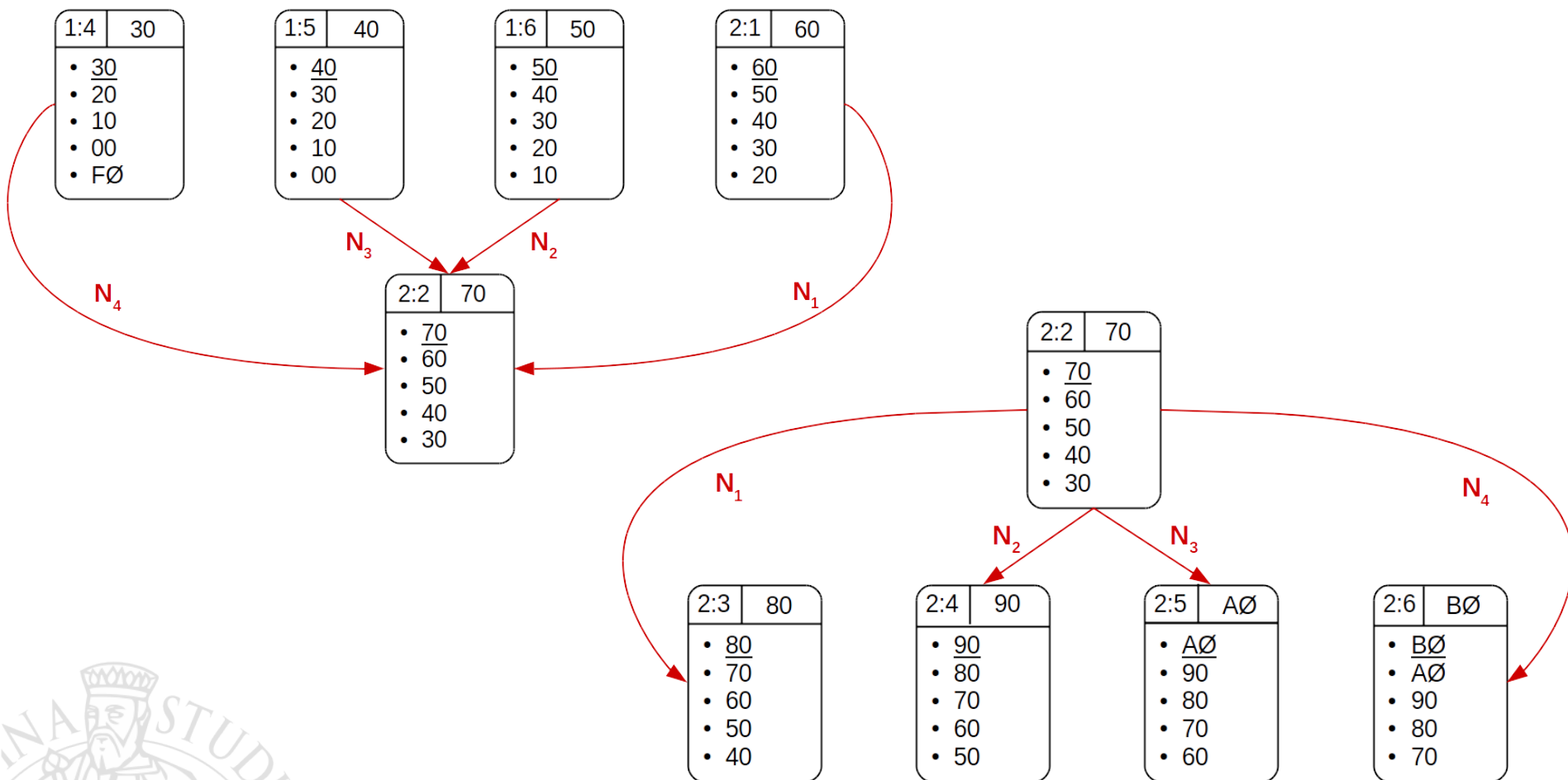
2a01:84a0:1001:a001::2:4;90;a0;90

2a01:84a0:1001:a001::2:5;a0;b0;a0

2a01:84a0:1001:a001::2:6;b0;c0;b0



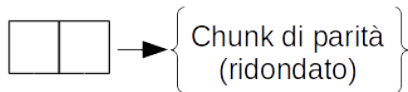
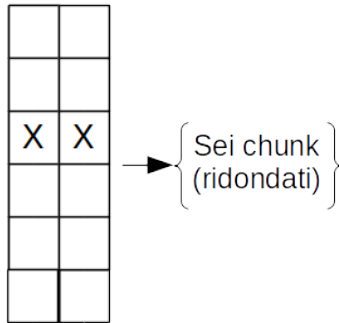
Sottoscrizioni native



Replica dato

- ID chunk e contenuto
- Funzione restituisce IP della board e il successivo
- Aperte due connessioni parallele
- Query in parallelo che permette la scrittura dei chunk su due host differenti
- Chiusura connessioni

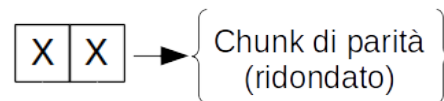
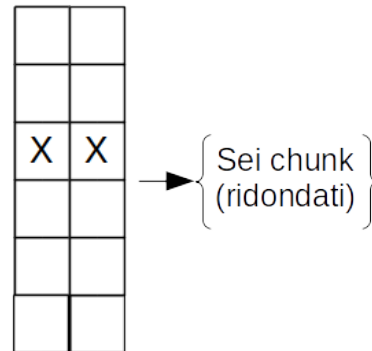
Replica dato



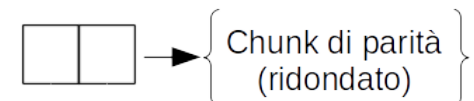
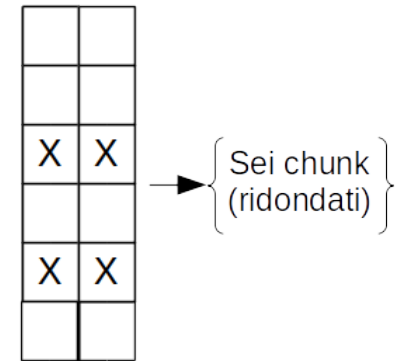
Perdita di un chunk

Quanti dischi possono essere danneggiati per causare la perdita di un dato?

Perdita di un chunk e la sua parità

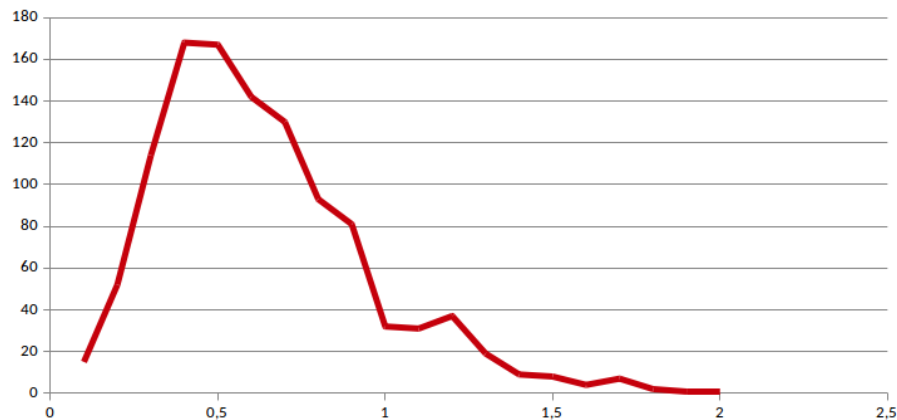


Perdita di due chunk



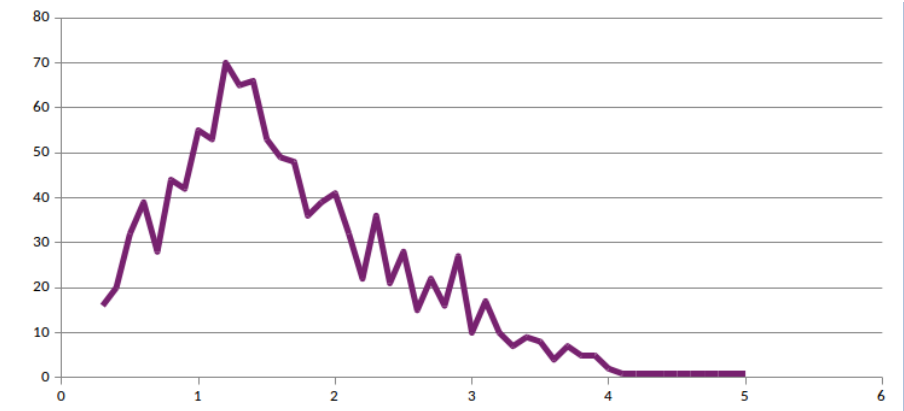
Replica dato

Ascisse: tempo impiegato per ogni dato

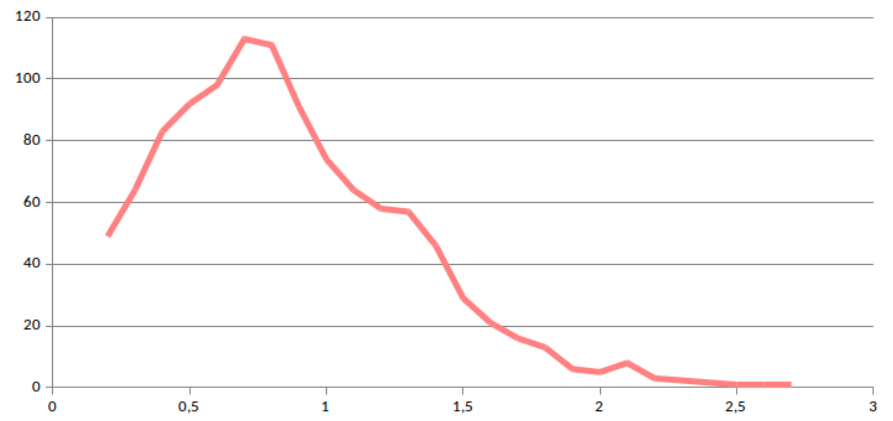


Nessuna copia di ridondanza

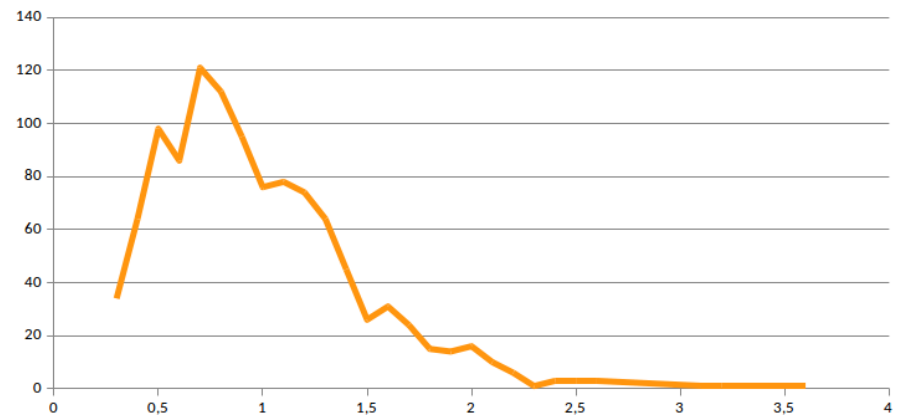
Ordinate: frequenza di PUT registrate



Connessioni e PUT sequenziali,
sottoscrizioni sincrone

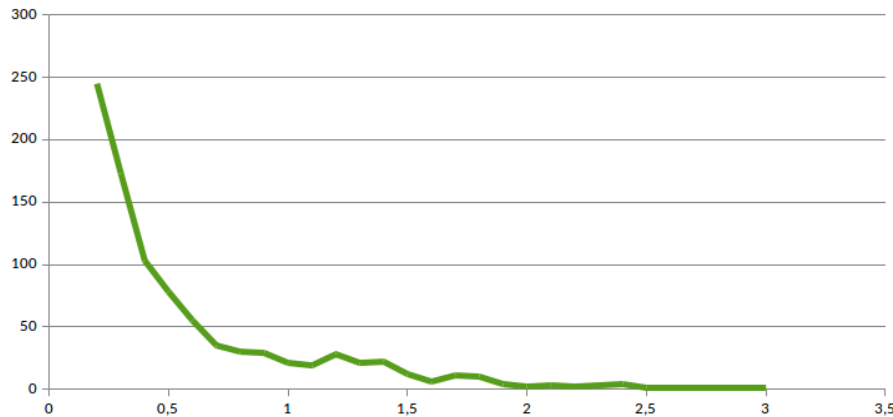


Sottoscrizioni asincrone

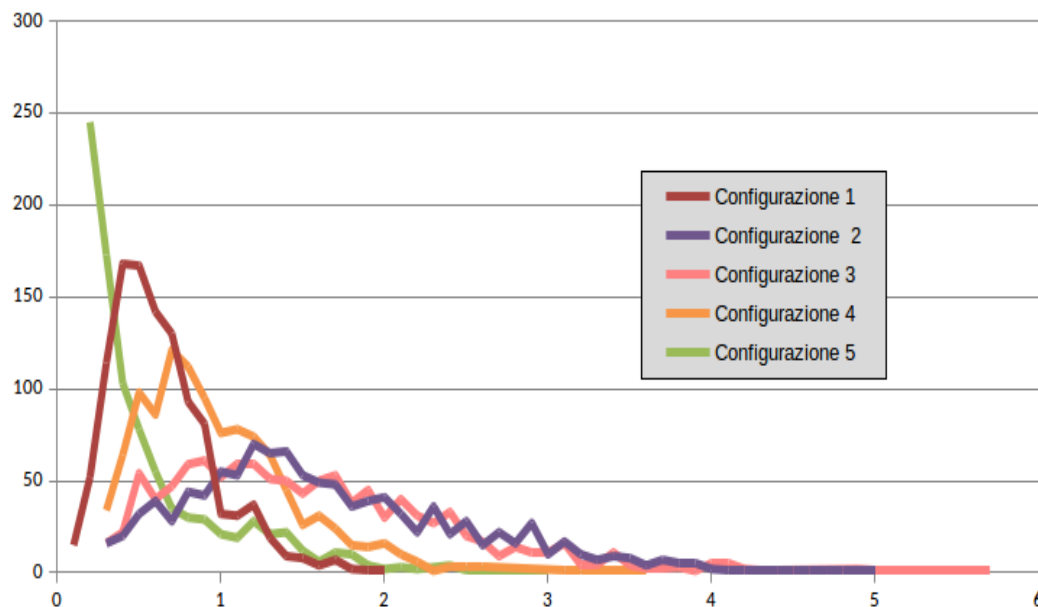


Connessioni sequenziali e PUT in parallelo

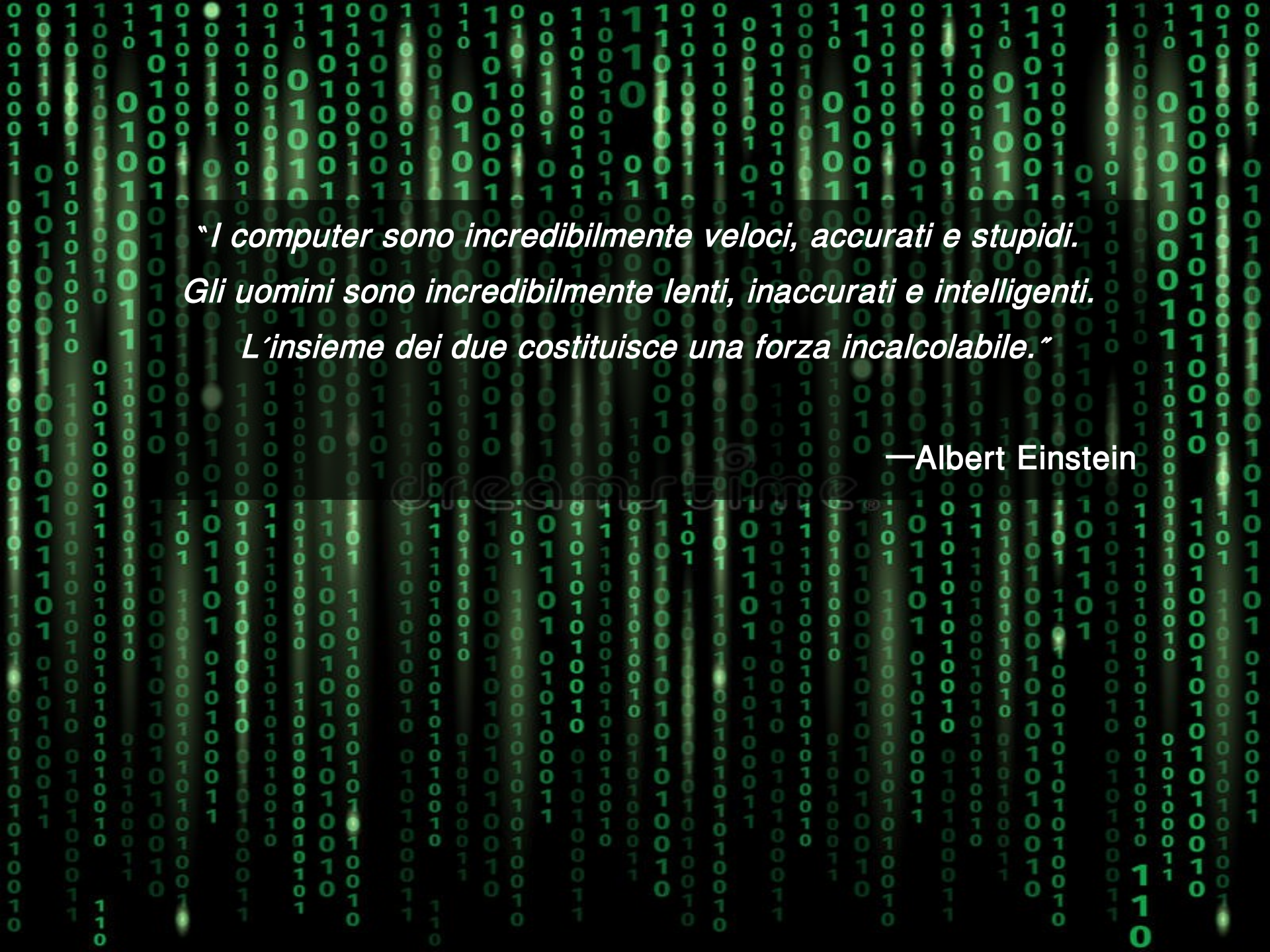
Replica dato



- Connessioni aperte in parallelo
- Scrittura sul database, senza attendere messaggi di successo
- Compromesso con il caso ottimale



- Evoluzioni hardware
 - Utilizzo di processori dual core
 - Utilizzo di dischi SSD



*“I computer sono incredibilmente veloci, accurati e stupidi.
Gli uomini sono incredibilmente lenti, inaccurati e intelligenti.
L'insieme dei due costituisce una forza incalcolabile.”*

—Albert Einstein