

RAPPORTI TECNICI INGV

Architettura e realizzazione
dell'infrastruttura tecnologica del
Centro di Monitoraggio delle
Attività del Sottosuolo



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

452

Direttore Responsabile

Valeria DE PAOLA

Editorial Board

Luigi CUCCI - Editor in Chief (luigi.cucci@ingv.it)
Raffaele AZZARO (raffaele.azzaro@ingv.it)
Christian BIGNAMI (christian.bignami@ingv.it)
Viviana CASTELLI (viviana.castelli@ingv.it)
Rosa Anna CORSARO (rosanna.corsaro@ingv.it)
Domenico DI MAURO (domenico.dimauro@ingv.it)
Mauro DI VITO (mauro.divito@ingv.it)
Marcello LIOTTA (marcello.liotta@ingv.it)
Mario MATTIA (mario.mattia@ingv.it)
Milena MORETTI (milena.moretti@ingv.it)
Nicola PAGLIUCA (nicola.pagliuca@ingv.it)
Umberto SCIACCA (umberto.sciacca@ingv.it)
Alessandro SETTIMI (alessandro.settimi1@istruzione.it)
Andrea TERTULLIANI (andrea.tertulliani@ingv.it)

Segreteria di Redazione

Francesca DI STEFANO - Coordinatore
Rossella CELI
Robert MIGLIAZZA
Barbara ANGIONI
Massimiliano CASCONI
Patrizia PANTANI
Tel. +39 06 51860068
redazione@ingv.it

REGISTRAZIONE AL TRIBUNALE DI ROMA N.174 | 2014, 23 LUGLIO

© 2014 INGV Istituto Nazionale
di Geofisica e Vulcanologia
Rappresentante legale: Carlo DOGLIONI
Sede: Via di Vigna Murata, 605 | Roma



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

RAPPORTI TECNICI INGV

Architettura e realizzazione
dell'infrastruttura tecnologica del Centro di
Monitoraggio delle attività del Sottosuolo

*Architecture and realization of the technological
infrastructure of the CMS activities*

Giampaolo Zerbinato^{1,3}, Alexander Garcia^{1,3}, Daniela Famiani^{2,3}, Thomas Braun^{2,3}

¹INGV | Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna

²INGV | Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Sismologia Tettonofisica

³INGV | Centro di Monitoraggio delle attività del Sottosuolo (CMS)

Accettato 21 settembre 2021 | Accepted 21 September 2021

Come citare | How to cite Zerbinato G. et al., (2022). Architettura e realizzazione dell'infrastruttura tecnologica del Centro di Monitoraggio delle attività del Sottosuolo. Rapp. Tec. INGV, 452: 1-32, <https://doi.org/10.13127/rpt/452>

In copertina Esempio di localizzazione | Cover Localization example

452

INDICE

Riassunto	7
<i>Abstract</i>	7
Introduzione	7
1. Infrastruttura informatica: Analisi delle possibili soluzioni	8
1.1 Infrastruttura a 2+2 macchine su 2 hypervisor	8
1.2 Dimensionamento dischi	10
1.2.1. Valutazione soluzioni: Ipotesi 1	11
1.2.2. Valutazione soluzioni: Ipotesi 2	12
1.3. Posizione del database	13
1.4. Posizionamento fisico	13
2. Acquisizione e processing automatico	14
2.1 Perché questa scelta?	14
2.1.1 Altre soluzioni	15
2.1.2 SeisComP Data Structure (SDS)	15
2.2 Protocolli di acquisizione	15
2.2.1 SeedLink	15
2.2.2 CAPS	16
2.2.3 SeedLink vs CAPS	16
2.3 Estrazione dei dati dall'archive creato tramite CAPS	17
2.3.1 Il formato miniSEED	18
2.3.2 Analisi del formato miniSEED e proposta di soluzione per la conversione in miniSEED	18
2.4 Pipeline processo di analisi	21
2.5 Flussi di dati	23
2.5.1 Flusso complessivo dei dati ricevuti	23
2.5.2 Flussi dati provenienti dai singoli canali	23
3. Messaggistica GDS	25
4. Conclusioni	26
Bibliografia	26
Sitografia	26

Riassunto

In questo articolo viene descritta l'infrastruttura informatica implementata dal Centro per l'acquisizione ed elaborazione dei dati utili al monitoraggio sismico e geodetico delle attività di Sottosuolo nelle aree di competenza di cui, secondo quanto disposto dagli Indirizzi e Linee Guida ed in seguito alla nomina da parte del Ministero per lo Sviluppo Economico, l'INGV risulta Struttura Preposta al Monitoraggio. Questo articolo è focalizzato in particolare sulla descrizione dell'infrastruttura *hardware*, *software*, la loro installazione e configurazione di base, ed i formati di dati usati.

Abstract

This article describes the IT infrastructure implemented by the Centre for the Monitoring of Subsoil activities to monitor the areas of competence of which, according to the provisions of the Addresses and Guidelines and following the appointment by the Ministry for Economic Development, the INGV is the Structure in charge of seismic and geodetic Monitoring. Particular attention is paid to the hardware and software infrastructure, the data formats used and their installation is described.

Keywords Monitoraggio tempo reale | Real time monitoring.

Introduzione

Il Centro per il Monitoraggio delle attività di Sottosuolo (CMS) è la struttura dell'Istituto Nazionale di Geofisica e Vulcanologia (INGV) dedicata all'applicazione delle Linee Guida proposta dal Ministero dello Sviluppo Economico per il monitoraggio della sismicità delle deformazioni del suolo e delle pressioni di poro. Il CMS ha il compito della raccolta, analisi, interpretazione e modellazione dei dati sismici e geodetici relativi alle aree del territorio nazionale interessate da attività antropiche di sfruttamento di *georisorse* (quali estrazione di idrocarburi, *reiniezione* di acque di scarto, stoccaggio di gas naturale, *geotermia*, *sequestro* di anidride carbonica, estrazioni minerarie, invasi artificiali) o legate ad altri rischi antropogenici [CMS, 2021]. Il CMS usa una infrastruttura ridondante per acquisire e processare i dati sismologici di sua competenza. Questa infrastruttura è descritta in questo articolo.

Il CMS all'atto della sua fondazione aveva, tra i primi obiettivi da realizzare per l'avvio delle sue attività, la costruzione di una infrastruttura tecnologica che rispondesse a requisiti di affidabilità, alta disponibilità e performance, resilienza e sicurezza. Le necessità primarie erano di conservare sia dati vincolati da accordi di riservatezza sia dati pubblici, di acquisirli senza interruzione 24 ore su 24, 7 giorni su 7, e di permettere agli operatori di eseguire le necessarie elaborazioni in qualsiasi momento e da qualsiasi luogo, senza vincolo alcuno di orario. Per rispondere a tali richieste è stato disegnato e implementato il prototipo di infrastruttura che descriviamo nei prossimi paragrafi.

1. Infrastruttura informatica: Analisi delle possibili soluzioni

La disponibilità, la confidenzialità e la sicurezza dei dati sono da sempre principi cardine alla base della realizzazione di infrastrutture informatiche. Uno degli obiettivi principali definiti da parte del direttivo del CMS è stato configurare l'infrastruttura in maniera ridondante, in modo tale da ridurre al massimo i rischi di perdita di dati e massimizzare la sicurezza dell'informazione relativa alle attività di monitoraggio. Per l'implementazione dell'infrastruttura sono stati presi in considerazione diversi prototipi valutando tutte le caratteristiche considerate idonee e non idonee agli scopi del CMS. Ed è proprio a seguito di considerare le particolarità che contraddistinguono le attività del CMS che il personale tecnico-scientifico afferente al CMS si è adoperato per attuare una soluzione che soddisfacesse la richiesta di protezione dei dati e di affidabilità del sistema.

Per soddisfare i requisiti di alta disponibilità ed affidabilità sono state prese in considerazione varie alternative. La soluzione canonica di un'unica macchina in cui i compiti di acquisizione, elaborazione e archiviazione fossero in essa racchiusi è stata esclusa a priori. Il vantaggio principale di una tale implementazione sarebbe stato la rapidità con cui questa configurazione sarebbe stata implementata e resa operativa, ovviamente a scapito di tutte le altre prerogative di cui si è discusso sopra.

Poi si è passati a considerare la virtualizzazione. Le macchine virtuali hanno, tra i vari vantaggi, quello di poterne salvare lo stato prima di eseguire operazioni che, se fallissero, potrebbero comprometterne il funzionamento. Inoltre, all'occorrenza, le macchine virtuali possono essere facilmente spostate da un server fisico ad un altro. Deciso che la via della virtualizzazione era quella da seguire, ci si è dunque orientati verso una topologia infrastrutturale distribuita. Dati i vincoli operativi, abbiamo considerato più efficiente la soluzione che prevede l'implementazione di due macchine gemelle che acquisiscano ed elaborino dati in modo indipendente l'una dall'altra. In questa configurazione una singola macchina si occuperebbe di acquisire, archiviare ed elaborare il dato sismologico ricevuto. Il vantaggio evidente di questa configurazione si basa sull'aumento dell'affidabilità del sistema; ad esempio, se una macchina ha problemi tecnici l'altra può continuare a funzionare indipendentemente. Lo svantaggio di questa configurazione consiste nella complessità di gestione e sincronizzazione dei risultati dell'elaborazione dei dati. Questo ultimo punto ha reso necessario effettuare ulteriori considerazioni per l'architettura del sistema, di seguito esposte.

La prima ipotesi considerata è una struttura con due coppie di macchine virtuali, realizzate su due macchine fisiche identiche (i.e., quattro macchine virtuali in totale). Ogni coppia di macchine sarebbe stata costituita da una di acquisizione e archiviazione e da un'altra di *processing*, *database storage* e di post-processing dei dati. Con questa configurazione, se una funzione venisse a mancare, si avrebbe solo il 25% dell'infrastruttura compromessa. Quindi, pur essendo la funzionalità compromessa per il 50%, la diagnostica e la riparazione riguarderebbe solo il 25% dell'infrastruttura. È stato quindi scelto di implementare questa soluzione, come viene dettagliatamente indicato nel paragrafo successivo.

1.1 Infrastruttura a 2+2 macchine su 2 hypervisor

Tenendo conto delle ipotesi valutate e descritte nel paragrafo precedente, per l'implementazione dell'infrastruttura di acquisizione e analisi dati del CMS si è scelto di utilizzare due server fisici gemelli, che ospitano due macchine virtuali ciascuno, che operano in parallelo ed indipendentemente l'uno dall'altro. Ogni *server* fisico è costituito da una macchina virtuale che gestisce le attività di acquisizione dei dati in tempo reale e di una seconda macchina virtuale che si occupa dei compiti di elaborazione automatica dei dati, *storage* del *database* dei dati

elaborati, e inoltre gestisce le attività di *post-processing* fatta dagli analisti. La struttura modulare è stata scelta a scopi di *fault tolerance*. In caso di rottura di un server fisico, l'altro continua ad acquisire e processare, senza compromettere completamente l'operatività del sistema. Per quanto riguarda l'hardware installato, i due server fisici sono costituiti da *computer* con *cpu xeon* a 40 *core* in *hyper-threading*, e 4 TB di *ram* ciascuno. Le 4 macchine virtuali sono equipaggiate con 8 *core*, 20 GB di *ram*, ospitano il sistema operativo Linux Centos 8 e acquisiscono e processano i dati mediante *software* SeisComP [Seiscomp, 2021]. L'acquisizione dei dati sismici in tempo reale è eseguita utilizzando un modulo di "Common Acquisition Protocol Server" (CAPS) [CAPS, 2021]. I dati sismologici, i *database* e i volumi delle macchine virtuali sono conservati in uno *storage array* Dell scv-3000. Il diagramma della configurazione implementata è mostrato nella Figura 1.

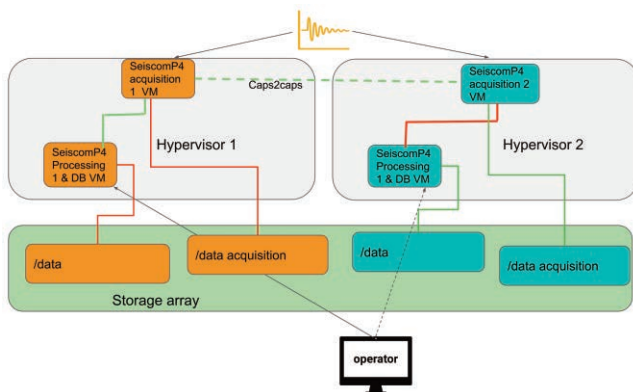


Figura 1 Configurazione dell'infrastruttura informatica di monitoraggio del CMS.

Figure 1 Configuration of the IT monitoring infrastructure of the CMS.

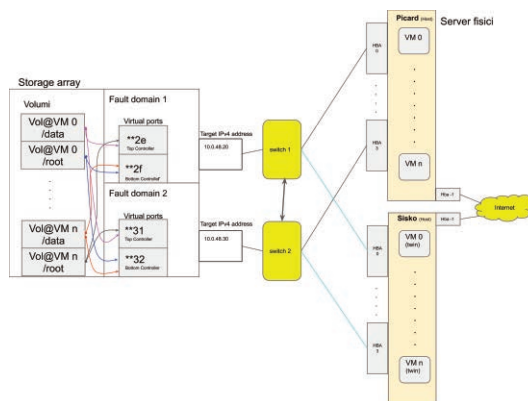
Il *software* di acquisizione e *processing* è stato installato su due macchine virtuali (Figura 1), la macchina *acquisition 1 VM* si occupa di acquisire e immagazzinare il dato sismologico proveniente dalle stazioni sismiche soggette al monitoraggio del CMS. Il dato viene immagazzinato in un *archive* gestito dal CAPS server. La macchina *processing 1 VM* processa il dato sismologico acquisito e reso disponibile dalla macchina *acquisition 1 VM*, entrambe installate sulla stessa unità fisica di hardware (*hypervisor1*). Entrambe le macchine virtuali ospitano il sistema operativo Centos 8 e sono entrambe equipaggiate con 2 volumi situati fisicamente nello *storage array*. Un volume è dedicato al sistema operativo e l'altro ai dati. Entrambe le macchine virtuali sono realizzate mediante *libvirt* [Libvirt, 2021] installata su sistema operativo Debian 10 sull'*hypervisor1*. La connessione tra l'*hypervisor* e lo *storage array* (Figura 1) è stata ottenuta mediante *device mapper multipath* [Device Mapper Multipathing, 2021].

Questa struttura è completamente gemellata sul *hypervisor2*, identico al precedente sia dal punto di vista del hardware che del software. L'*hypervisor2* ospita quindi le altre 2 macchine virtuali, *acquisition2* e *processing2*, anch'esse identiche alle rispettive macchine implementate sull'*hypervisor1*.

Le due coppie di macchine (ogni coppia sul proprio *hypervisor* fisico) lavorano contemporaneamente ed indipendentemente, così che se una macchina di una coppia dovesse fermarsi, l'altra coppia di macchine funzionerebbe regolarmente, permettendo il regolare svolgimento delle funzioni di acquisizione e *processing* del dato sismologico. Questa configurazione ha lo scopo di permettere una manutenzione scalata e di apportare modifiche al sistema, senza che si abbia la benché minima interruzione del servizio. Inoltre, i 2 *hypervisor* si collegano allo *storage array* mediante 2 schede *ethernet* in *bonding active-backup* [Bonding, 2021], le quali passano attraverso due distinti *switch* Netgear XS748T.

Figura 2 Schema di principio: come sono disposte le macchine fisiche e virtuali e come si collegano allo *storage array*.

Figure 2 Principal scheme: how the physical and virtual machines are arranged and how they connect to the *storage array*.



1.2 Dimensionamento dischi

A seguito di un'accurata analisi, riportata nel paragrafo 2.5 in merito al flusso dei dati sismologici in ingresso, è stato stimato che, una volta a regime, il sistema di acquisizione riceverà un quantitativo di dati pari a 13 GB al giorno. Secondo quanto indicato dal produttore Dell, e confermato euristicamente dall'esperienza, lo *storage array* scv-3000 in uso al CMS lavora in condizioni ottimali se non si supera circa l'80% della sua capacità nominale lorda. Infatti lo *storage array* Dell scv-3000 usa contemporaneamente lo schema di *raid* 10DM e *raid* 6-10 [Dell Engineering, 2016] in maniera del tutto autonoma, in cui non è prevista dalla casa costruttrice possibilità di intervento da parte dell'utente.

Lo schema di *raid* 10-DM presenta la caratteristica secondo cui il dato è scritto simultaneamente su 3 dischi separati, col vantaggio di maggiore performance in lettura/scrittura e di tolleranza ai guasti.

Mentre lo schema *raid* 6-10 usa 8 segmenti di dato e 2 segmenti di parità per ogni *stripe*, permettendo di verificare l'integrità del dato e di ricostruirlo in caso di guasto ad un disco. È meno performante rispetto al *raid* 10-DM.

Lo *storage array* Dell utilizza una policy di movimento dei dati, detto *data progression*, tra i 2 schemi di *raid* sopracitati, dove lo spostamento automatico dei dati tra i 2 schemi di *raid* punta a tenere il più possibile libero lo schema *raid* 10-DM a scapito del *raid* 6-10, con lo scopo di rendere disponibile il più performante *raid* 10-DM ai dati in uso alle applicazioni nel momento dell'esecuzione delle stesse. Il costruttore implementa questa politica, in maniera non controllabile dall'utente, perché, sostiene sempre il costruttore, statisticamente, solo il 20-30% circa dei dati contenuti nello *storage array* viene usato in maniera attiva dalle applicazioni, mentre il 70 all'80% è meno attivo. Il movimento dei dati è dinamico, e la visibilità delle allocazioni a livello di *raid* effettive può essere osservata dalla *dashboard* dello *storage array* stesso [Dell Engineering, 2016]. Quanto scritto sopra, tradotto in numeri, indica quanto segue:

			Caso peggiore	Caso migliore	
Size disk (TB)	n. disks	Spazio disponibile lordo	Se tutto lo spazio fosse immagazzinato in <i>raid</i> 10 DM	Se tutto lo spazio fosse immagazzinato in <i>raid</i> 6-10	Livello di occupazione da non superare
7.28	16	116.48	38.83	93.18	80TB

Tabella 1 Valutazione degli scenari migliore e peggiore negli schemi di *raid* dello *storage array* Dell scv-3000.
Table 1 Best- and worst-case evaluation in Dell scv-3000 *storage array* *raid* schemes.

Quindi, riassumendo, il volume della macchina di acquisizione non può essere troppo grande, altrimenti lo *storage array* rischia di non funzionare più correttamente. Basandosi sul flusso di dati allo stato attuale (i.e., non a pieno regime) pari a 5,7 GB al giorno, abbiamo ipotizzato due scenari di crescita del flusso di dati (Ipotesi 1 e Ipotesi 2).

1.2.1 Valutazione soluzioni: Ipotesi 1

Considerando un periodo di transizione fino a regime di 18 mesi, ed una crescita del flusso di dati come mostrato nella Figura 3, il flusso di dati giornaliero e la stima cumulata dei dati raccolti nel tempo avrebbe un andamento come quello mostrato nella Tabella 2.

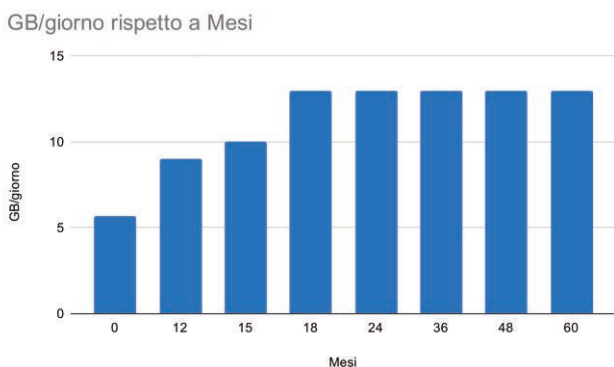


Figura 3 Flusso dei dati giornaliero.
Figure 3 Daily data flow.

Inizio periodo (mese)	Fine periodo (mese)	Flusso di dati (GB/giorno)	Dati totali raccolti solamente nel periodo considerato (GB)	Dati totali raccolti dall'inizio fino al termine del periodo considerato (GB)
0	12	5.7	2.052	2.052
13	15	9	810	2.862
16	18	10	900	3.762
19	24	13	2.340	6.102
25	36	13	4.680	10.782
37	48	13	4.680	15.462
49	60	13	4.680	20.142

Tabella 2 Flusso dei dati giornaliero secondo l'ipotesi 1 e valore totale dei dati memorizzati al termine di ogni periodo, calcolato secondo la seguente formula, dove per *Numero_giorni_{medio mensile}* si è preso il valore di 30 giorni.

$$Dati\ totali\ raccolti_{nel\ periodo\ considerato} = Flusso_{dati\ giornaliero} * (Fine_{periodo} - Inizio_{periodo}) * Numero_{giorni}_{medio\ mensile}$$

Table 2 Daily data flow according to hypothesis 1 and total value of the data stored at the end of each period of time, calculated according to the following formula, where for *Number days_{monthly average}* the value of 30 days was taken.

$$Total\ data\ collected_{in\ the\ considered\ period} = Flow_{daily\ data} * (End_{period} - Start_{period}) * Number\ days_{monthly\ average}$$

In questa ipotesi un volume di 21 TB è sufficiente a coprire un periodo di tempo pari a 5 anni. Si osserva che 5 anni è il periodo di garanzia dell'infrastruttura informatica attualmente

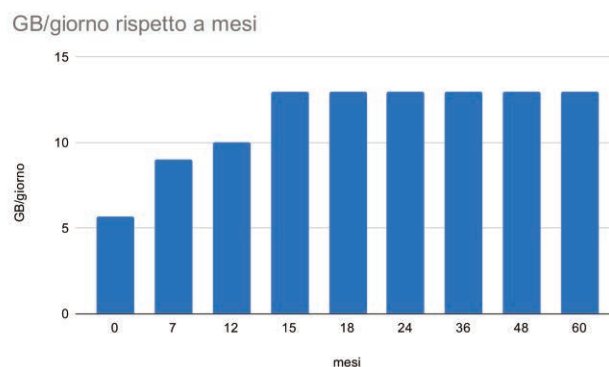
in possesso al CMS. Dopo tale periodo è più che ragionevole affermare la necessità di una sua sostituzione.

1.2.2 Valutazione soluzioni: Ipotesi 2

Ipotizzando un periodo di transizione fino a regime di 15 mesi, con la crescita del flusso di dati riportato nella Figura 4, il flusso di dati giornaliero e la stima cumulata dei dati raccolti nel tempo avrebbe un andamento come quello mostrato nella Tabella 3.

Figura 4 Flusso dati giornaliero nei vari mesi.

Figure 4 Daily data flow in the various months.



Inizio periodo (mese)	Fine periodo (mese)	Flusso di dati (GB/giorno)	Dati totali raccolti nel periodo considerato (GB)	Dati totali raccolti dall'inizio fino al termine del periodo considerato (GB)
0	7	5.7	1.197	1.197
8	12	9	1.350	2.547
13	15	10	900	3.447
16	18	13	1.170	4.617
19	24	13	2.340	6.957
25	36	13	4.680	11.637
37	48	13	4.680	16.317
49	60	13	4.680	20.997

Tabella 3 Flusso dei dati giornaliero secondo l'ipotesi 2 e valore totale dei dati memorizzati al termine di ogni periodo, calcolato secondo la seguente formula, dove per $Numero_giorni_{medio\ mensile}$ si è preso il valore di 30 giorni.

$$Dati\ totali\ raccolti_{nel\ periodo\ considerato} = Flusso_{dati\ giornaliero} * (Fine_{periodo} - Inizio_{periodo}) * Numero_giorni_{medio\ mensile}$$

Table 3 Daily data flow according to hypothesis 2 and total value of the data stored at the end of each period of time, calculated according to the following formula, where for $Number\ days_{monthly\ average}$ the value of 30 days was taken.

$$Total\ data\ collected_{in\ the\ considered\ period} = Flow_{daily\ data} * (End_{period} - Start_{period}) * Number\ days_{monthly\ average}$$

Tenendo conto delle stime mostrate secondo le due ipotesi precedenti, si considera che un volume con capacità di storage di 22TB, 1TB più grande dello scenario peggiore, è in ogni caso

sufficiente a coprire un periodo di tempo di circa 5 anni. Inoltre, usando queste stime è possibile valutare la percentuale di riempimento dello *storage array* al termine del suo presunto periodo di funzionamento, come mostrato nella Tabella 4.

Volume (TB)	Spazio corrispondente nello <i>storage array</i> (TB)	Occupazione rispetto al limite di sicurezza dello <i>storage array</i> (80TB)	Spazio utile rimanente nello <i>storage array</i> (TB)
21	42	52,50%	38
22	44	55,00%	36

Tabella 4 Confronto tra le ipotesi 1 e 2 sulla velocità di riempimento dello *storage array*.
Table 4 Comparison between the hypotheses 1 and 2 on the *storage array* filling speed.

L'intera valutazione dello *storage* necessario fin qui esposta non tiene conto del fatto che generalmente tutti i *datalogger* comprimono i dati in compressione *Steim1* o *Steim2*. Il rapporto di compressione varia tra 2.5 e 3 a seconda della rumorosità del sito e del livello di sismicità. Perciò tutte le considerazioni fatte fin qui sono da considerare estremamente conservative poiché l'occupazione effettiva sarà inferiore rispetto alle previsioni (vedi §2.5).

Si può pertanto concludere che entrambi i dimensionamenti sono compatibili con la capacità dello *storage array*, per un orizzonte temporale di 5 anni. Si osserva che, con volumi di dati di tale dimensione, per spostare un volume è necessaria una espansione, detta *enclosure*, oppure l'utilizzo di un nuovo *storage array*.

In conclusione sono stati creati volumi da 22 TB. Con questo dimensionamento, anche nel caso peggiore, rappresentato da un uso eccessivo dello schema *raid 10 DM* da parte dello *storage array*, è possibile prefigurare una capacità operativa di *storage* non inferiore ai 5 anni, fermo restando il flusso di dati previsto nelle sezioni precedenti.

1.3 Posizione del database

Tutti i risultati dell'elaborazione dei dati effettuata dalla macchina di *processing* sono memorizzati in un database MySQL. Un test durato soli sei mesi e con flusso di dati ridotto (5-6 GB/giorno) ha prodotto un *database* di 12 GB. Abbiamo pertanto deciso, vista la veloce crescita del *database*, di installarlo in un apposito volume. Si ricorda che uno *storage array* sono tanti dischi fissi separati che vengono uniti logicamente. Per renderlo fruibile a più macchine, questo viene partizionato, ogni partizione si chiama volume, ed ha le dimensioni che vengono decise dal sistemista in fase di creazione. Andando maggiormente nel dettaglio, il disco è l'oggetto fisico (*hdd, ssd, penna usb, floppy disk*), la partizione è una parte di disco, mentre il volume è una partizione con il *filesystem* [maketecheasier.com, 2021].

1.4 Posizionamento fisico

Come previsto dallo schema di trasmissione ed acquisizione, i dati trasmessi dai server di acquisizione dei concessionari fornitori dei dati arrivano in formato miniSEED. Presso il Centro acquisizione dati della sezione INGV di Bologna, CAD-BO, tali dati vengono ri-acquisiti avvalendosi di un protocollo di acquisizione commerciale, chiamato *Common Acquisition Protocol*

Server (CAPS), il cui vantaggio principale, rispetto al più diffuso SeedLink Server, consiste nella possibilità di riconfigurarli in corso d'opera, senza interruzione e successiva ripartenza del sistema.

2. Acquisizione e processing automatico

SeisComP è una raccolta di *software* per l'acquisizione di dati sismologici, l'analisi e il monitoraggio in tempo reale di terremoti e altri eventi naturali o antropici che generano dei segnali potenzialmente localizzabili, attraverso l'elaborazione interattiva e lo scambio di dati su *Internet*. È stato sviluppato da GeoForschungsZentrum (GFZ) originariamente per la rete sismologica GEOFON. A seguito del terremoto M9.2 a Sumatra (2004), nell'ambito del progetto GITEWS [Gitews, 2021], tale *software* è stato modernizzato implementando l'interfaccia grafica e aggiungendo numerosi strumenti per l'analisi dati.

Nel 2020 è stata pubblicata la versione SeisComP4 (versione utilizzata dal CMS per le sue attività di monitoraggio nel momento in cui si scrive), che è *public domain* e *open source*; continuano a rimanere a pagamento (con licenza) i moduli aggiuntivi gestiti da Gempa GmbH.

2.1 Perché questa scelta?

SeisComP consta di due versioni, una versione gratuita ed una a pagamento chiamata SeisComPro. SeisComP, la versione gratuita [SeisComp3, 2021], include strumenti di elaborazione per il rilevamento degli eventi.

I suoi moduli principali sono:

- *seedlink*: adibito all'acquisizione e l'archiviazione dei segnali sismici;
- *scautopick*, che trova anomalie sui segnali continui in termini di ampiezza (anche chiamati *picks*);
- *scautoloc*, che identifica le combinazioni di *picks* che corrispondono a un evento sismico comune e ne determina l'ipocentro;
- *scamp* / *scmag*, per calcolare la magnitudo;
- *scevent*, che associa uno o più ipocentri localizzati ad un unico evento.

SeisComPro [SeisComPro, 2021] estende SeisComP *free* con i seguenti moduli:

- *scanloc* adibito ad individuare le fasi P e S;
- *ccloc* per rilevare e discriminare eventi applicando la correlazione incrociata basata su modelli di forme d'onda;
- *sceval* per discriminare eventi sismici reali da eventi errati o falsi che possono verificarsi durante il monitoraggio utilizzando soglie molto basse per il rilevamento delle anomalie d'ampiezza;
- *GAPS* per accedere al sistema SeisComP da qualsiasi sistema ovunque sulla Terra tramite un *browser web*;
- *CAPS* è l'alternativa al *server* di acquisizione di SeedLink/ArcLink. Consente la ridondanza nel processo di acquisizione e la variazione della configurazione delle reti sismiche associate al sistema, senza interrompere l'intero processo di acquisizione ed analisi;
- *npeval* un modulo per valutare le prestazioni della rete in tempo reale e per ottimizzare i progetti di rete;
- *scqceval* per riconfigurare dinamicamente le reti in base ai parametri di qualità della forma d'onda (Quality Control).

I punti di forza che hanno convinto il CMS ad adottare SeisComPro, la versione professionale, come *software* per il monitoraggio sono la grande versatilità del sistema di acquisizione e

archiviazione dei dati sismici che propone, la larga diffusione del software e il contributo continuo di una vasta comunità di sismologi che usano il software e ne individuano, con la pratica continua, eventuali imperfezioni ed errori. Inoltre, la presenza di interfacce grafiche già integrate al sistema facilitano l'analisi interattiva dei risultati [Olivieri and Clinton, 2012; Braun et al., 2020; Braun et al., 2019; Danesi et al., 2020].

2.1.1 Altre soluzioni

Esistono altri *software* realizzati per le attività di monitoraggio sismico e vengono di seguito elencati. Il *software* Earthworm [EarthwormCentral, 2021] è un pacchetto *software* scritto in C, sviluppato da oltre 25 anni da United States Geological Survey (USGS). Lavora su ogni piattaforma, Windows incluso. Può essere usato anche nell'ambito di onde infrasuono, geomagnetiche e per misurazioni atmosferiche [Earthworm, 2021]. Questo sistema di acquisizione dei dati sismici viene utilizzato nella "Sala di Monitoraggio Sismico e Tsunami" dell'Osservatorio Nazionale dei Terremoti (ONT). È un *software open source*, distribuito dall'USGS con una vasta comunità di utilizzatori e contributori; le nuove *release* sono gestite da ISTI, *Instrumental Software Technologies, Inc.*, che fornisce a pagamento supporto tecnico e corsi di formazione.

Un'altra soluzione può essere Antelope [Antelope, 2021]; anch'esso è nato per il monitoraggio ambientale, è fornito di una interfaccia documentata che permette lo sviluppo di *software* addizionali nei linguaggi C, Tcl/tk, Shell *script*, Fortran e Perl. È un *software* commerciale, venduto da Boulder Real Time Technologies, Inc. (*brtt.com*).

2.1.2 SeisComP Data Structure (SDS)

Il cuore di SeisComP è la *SeisComP Data Structure* [SDS, 2021] (SDS) 1.0 che è una struttura gerarchica di *file* e *directory* così costituita:

```
~/archive/YYYY/NW/STAT/CH/NW.STAT.LOC.CH.D.YYYY.JJJ
NW - Network code (codice rete)
STAT - Station code (codice stazione)
LOC - Location code (codice sito)
CH - Channel code (codice canale)
D - datastream type code (codice datastream)
YYYY - Year (anno)
JJJ - Julian day (giorno giuliano)
```

SDS è pensato come un paradigma per lo *standard* e la portabilità dei *server* di dati e del pacchetto di analisi.

2.2 Protocolli di acquisizione

2.2.1 SeedLink

Il protocollo SeedLink [SeedLink, 2021] è basato su una robusta trasmissione di dati destinata all'uso su Internet o circuiti privati che supportano TCP/IP. Il protocollo è robusto in quanto i *client* possono disconnettersi e riconnettersi senza perdere dati, in altre parole le trasmissioni possono essere riprese fintanto che i dati esistono ancora nel *buffer* del *server*. I flussi di dati

richiesti possono essere limitati a reti, stazioni, posizioni e/o canali specifici. Tutti i pacchetti di dati sono *record* miniSEED da 512 *byte*.

Il protocollo SeedLink può essere riassunto come una semplice fase di selezione dati basata su ASCII seguita dallo *streaming* di pacchetti di dati dal *server*. I pacchetti SeedLink sono composti da una piccola intestazione seguita da un *record* miniSEED da 512 *byte* (solo dati SEED). La fase di negoziazione consente al *client* di richiedere solo i dati specificati dal *server* per ogni flusso di dati selezionato. Un flusso di dati è definito da una coppia di codici di rete e stazione. I miglioramenti al protocollo SeedLink supportato dal *server* IRIS DMC (<https://ds.iris.edu/ds/nodes/dmc/about/facility/>) consentono di utilizzare caratteri jolly per i codici di rete e stazione oltre che per i codici di posizione e di canale.

Utilizzando i numeri di sequenza per ogni pacchetto in un flusso di dati, il protocollo SeedLink consente di riprendere le connessioni, eliminando la maggior parte delle lacune di dati. La capacità di riprendere i flussi di dati interrotti dipende principalmente dalla quantità di dati che, in termini di tempo, il SeedLink remoto ha nel suo *buffer*.

Il protocollo consente due diverse modalità di trasmissione dei dati, modalità *uni-station* e *multi-station*. La modalità *uni-station* funziona trasmettendo un singolo flusso di dati (dati da una singola stazione) attraverso una connessione di rete. In questa modalità il flusso di dati non deve essere specificato dal *client* poiché è implicito dall'indirizzo e dalla porta. La modalità *multi-station* funziona trasmettendo flussi di dati *multiplexati* (dati da più stazioni) attraverso una singola connessione di rete. Quasi tutti i collegamenti vengono negoziati come *multi-station*, anche se viene richiesta una sola stazione; la modalità *uni-station*, per la maggior parte dei *server* accessibili pubblicamente, è considerata obsoleta e sconsigliata [SeedLink, 2021]. SeedLink è stato originariamente creato GEOFON.

2.2.2 CAPS

Il protocollo CAPS [CAPS data management, 2021] nasce come una integrazione del protocollo SeedLink; in pratica, i dati sismologici nell'*archive* creato utilizzando CAPS hanno una struttura che, rispetto al formato *standard* miniSEED, differisce nell'aggiunta di un *header* [CAPS, 2021], secondo questo schema:

```
CAPS-format = SID-chunk + HEAD-chunk + miniSEED records
```

dove:

- SID-chunk è il campo dell'*header* ID;
- HEAD-chunk è il campo dell'*header* del *data information*;
- miniSEED record è il dato sismologico in formato miniSEED [MiniSEED, 2021].

CAPS aggiunge al *file* di dati miniSEED un SID di intestazione ID di flusso di dimensioni non fisse e un blocco di testa di 15 *byte* di dimensione fissa che precede i dati sismologici. Una trattazione più' dettagliata è fatta nella sezione "*Analisi del formato miniSEED e proposta di soluzione per la conversione in miniSEED*".

2.2.3 SeedLink vs CAPS

Tenendo conto dei requisiti operativi del CMS, che gestisce una sala sismica virtuale non supervisionata, l'affidabilità del sistema di acquisizione ed elaborazione dati è di fondamentale importanza per assicurare la continuità del servizio. In questo senso, la struttura del protocollo CAPS scelto per le operazioni di monitoraggio sismico del CMS ha diversi vantaggi rispetto

alla performance di SeedLink, come ad esempio:

- la possibilità di acquisire in parallelo e con un unico protocollo di acquisizione dati provenienti da sensori multiparametrici e i risultanti flussi di dati con un passo di campionamento differente (di interesse nel caso che in un sito siano collocati sensori *accelerometrici*, *velocimetrici*, CGPS, di temperatura, video camera, ecc.);
- variare (aggiungere o togliere) il numero di stazioni sismiche operative, senza dover interrompere l'acquisizione;
- sincronizzazione dei dati acquisiti attraverso i due *hypervisor* che operano indipendentemente, rendendo estremamente efficiente la ridondanza desiderata;
- *backfill option*: la possibilità, in caso di interruzione dell'acquisizione, o riavvio del *server*, di recuperare via ritrasmissione i *data-packages* persi e sincronizzarli automaticamente nell'archivio;
- controllo dell'accesso ai dati via autenticazione (*login/password*), nessun accesso pubblico ai dati;
- CAPS supporta servizi come *Winston Wave Server* [Winston, 2021] e *FDSN Wave Server* [FDSNWS, 2021].

La principale differenza tra SeisComP e SeisComPro è il *server* di acquisizione, rispettivamente SeedLink e CAPS. CAPS è un archiviatore di forme d'onda che consente l'acquisizione di diverse sorgenti di dati, come ad es. di temperatura, GPS, video e ovviamente dati sismici. Il *server* CAPS è inoltre caratterizzato dall'abilità di sincronizzare i suoi dati con altri *server* CAPS per creare una ridondanza dei dati sismologici. La peculiarità di CAPS è la possibilità di essere integrato con il componente CAPS2CAPS, un sincronizzatore di 2 archivi CAPS situati in 2 macchine diverse. Questa funzione non è disponibile in SeedLink. In pratica, l'operazione del servizio CAPS2CAPS permette alle due macchine, che stanno acquisendo in parallelo le stesse forme d'onda, di servire di supporto a vicenda: ad esempio, se una macchina si guasta, o viene fermata per manutenzione ordinaria, la seconda macchina continua ad acquisire. Quando la prima macchina torna in servizio, il modulo CAPS2CAPS ri-sincronizza i 2 archivi CAPS ospitati in entrambe le macchine. Per il tempo in cui la macchina 1 è ferma, l'operatore può utilizzare la seconda macchina e il lavoro non viene bloccato¹.

Il protocollo SeedLink ha però il vantaggio rispetto a CAPS di scrivere dati direttamente compatibili con moltissimi *tools* disponibili per le analisi di dati sismologici quali Sac [SAC, 2021] e ObsPy [ObsPy, 2021]. Per utilizzare tali *tools* con dati archiviati tramite CAPS è necessario estrarre il dato dall'archivio creato da CAPS tramite apposito software che scriva il dato sismologico nel formato *standard* miniSeed.

2.3 Estrazione dei dati dall'archive creato tramite CAPS

Come già menzionato sopra, archiviando il dato sismologico tramite protocollo CAPS si generano *file* scritti in formato miniSEED, ma con l'aggiunta di un *header* che rende il dato registrato nell'*archive* non direttamente fruibile tramite *tools* che utilizzano il formato *standard*. È possibile ottenere il dato dal CAPS *server* in formato *standard* tramite l'apposito *software* d'interrogazione dei dati Capstool [Capstool, 2021]. In questa sezione presentiamo una breve descrizione operativa del formato dei *file* e le differenze rispetto allo *standard* miniSEED.

¹ Questa non è certamente l'unica soluzione per implementare un'infrastruttura ad alta disponibilità; è possibile implementare strategie alternative come a esempio copiare l'archivio delle forme d'onda n volte al giorno in un altro volume di archiviazione oppure eseguire il backup in un'altra posizione, ma è più dispendioso in termini di tempo e risorse informatiche.

2.3.1 Il formato miniSEED

Il miniSEED è un sottoinsieme dello standard SEED utilizzato per serie temporali di dati. I *metadata* in questo formato includono una quantità limitata di informazione, come ad esempio l'identificazione delle serie temporali e semplici indicatori dello stato di salute. In particolare, nel *file* non sono inclusi dati come le coordinate geografiche della stazione, le informazioni di risposta strumentale e altre informazioni necessarie per interpretare i valori dei dati.

Un *file* o flusso di miniSEED è semplicemente una concatenazione di *record* di dati. A seconda delle capacità del *reader* considerato, i *record* di dati per più canali di dati possono essere *multiplexati* insieme. Esistono varie librerie di programmazione per supportare la facile lettura e scrittura di dati miniSEED senza conoscere i dettagli del formato, tra cui la libreria *libmseed* [MiniSEED, 2021].

2.3.2 Analisi del formato miniSEED e proposta di soluzione per la conversione in miniSEED

In questa sezione presentiamo un confronto dei *file* in formato miniSEED *standard* e quello scritto utilizzando il protocollo CAPS. Di seguito confrontiamo due *file* in formato binario dello stesso record sismologico, uno in formato *standard* e l'altro registrato tramite il protocollo CAPS, per individuarne le differenze. In primo luogo confrontiamo gli *header* iniziali di entrambi i *file* (Figure 5 e 6) successivamente confrontiamo le unità fondamentali del dato sismologico (Figure 7 e 8).

```

0 48454144 07000000 01000263 6E740044 41544100 02000031 35303038 30442053 HEAD cnt DATA 1500800 S
32 41525342 30304348 5A564407 E400E109 211E0000 A0019C00 32000500 00000200 ARSB00CHZVD % ! † ú Z
64 00000000 40003003 E800380A 01090003 E900005A 00000701 55555500 00015700 @ 0 È 8 È Z UUU W
96 000155FD 02090408 01050000 080102FB FCFCFF01 0006FE00 02040A03 0501FD02 U
128 00070C08 03FB06F4 F1F1EE7 EBEFF9F9 FBFF700 000C1115 55555514 1C0F1012 "0000A10..... UUU
160 150F06FD EFEF1F1 F0F6FAC 01010810 16140A04 02010006 1010110F 06FDF7EF "0000*..... UUU
192 E7E5E9F1 F6FE0503 02FD0300 FDFCF6F6 FCF0108 07070215 555555FE 030404FF AAÈ0..... UUU
224 FAF7FBFA F8F7F4F8 F8000709 0F050509 09080600 FFFCFDFD 02080706 05050607 ..... UUU
256 03DFBFBA 00040707 0401FCFC FEFCDFD FFE01FD F9FAF415 555555FB FDFC01FE ..... UUU
    
```

Figura 5 Header del file scritto tramite il protocollo CAPS.

Figure 5 Header in CAPS format.

```

0 31353030 38304420 53415253 42303043 485A5644 07E400E1 09211E00 00A0019C 1500800 SARSB00CHZVD % ! † ú
32 00320005 00000002 00000000 00400030 03E80038 0A010900 03E90000 5A000007 Z
64 01555555 00000157 00000155 FD020904 08010500 00080102 FBFCFCFF 010006FE UUU W U
96 0002040A 030501FD 02080F0C 0803FBF6 F4F1F1EE E7EBEFF9 F9FBFFF7 00000C11 UUU "0000A10..... UUU
128 15555555 141C0F10 12150F06 DFEFEF1 F1F0F6FA FC010108 1016140A 04020100 UUU "0000*..... UUU
160 06101011 0F06FDF7 EFE7E5E9 F1F6FE05 0302FD03 00FD0CF6 F6FCFF01 08070702 UUU "AAÈ0..... UUU
192 15555555 FE030404 FFAF7FB FAF8F7F4 F8F00007 090F0505 09090006 00FFCFD UUU ..... UUU
224 FD020807 06050506 0703FDFB FA000407 070401FC FCFEFCFD FDFFE01 FDF9FAF4 UUU ..... UUU
256 15555555 FBDFC01 FE01FFF0 02FF0001 FEFDFF01 0300FEFD FCFCF901 FCFD00F8 UUU ..... UUU
    
```

Figura 6 Header del file scritto in formato miniseed standard.

Figure 6 Header in standard miniseed format.

```

480 02FFFCFB FC080808 0C040808 0609FFF7 EFE9F0F3 FD020505 FEFF0207 0200FD5 ..... "0E*0"
512 F5FA0307 07FDF9F5 FC030400 03FEFF01 00E0E0E7 F8F2F444 41544100 02000031 ..... UUDATA
544 35303038 31442053 41525342 30304348 5A564407 E400E109 211F0019 F0019C00 50081D SARSB00CHZVD % ! † ú
576 32000500 00000200 00000000 40003003 E800380A 01090003 E900005A 00000701 Z
608 55555500 00015400 00014FFF 010402FC FA050809 0D0A00F7 F90200FF FDF4F2F3 UUU T O
640 F7FE0501 FE8F2FB 040D1818 0801FBFB 020304FF F2EEEA1F F8F9FCF8 F8FF1115 UUU "U
672 5555551C 191A09FF FFE0101 FEF1E8E5 F0F60007 07100E14 18141608 FAF3F0EB UUU ..... "0EA* "0E1
704 ECF1EEFZ F7FA0006 0F111512 000801FC F4F2EDF7 FFFC0104 0004090A 0804FA15 1000 ..... "0U1"
    
```

Figura 7 Unità fondamentale nel file scritto tramite protocollo CAPS.

Figure 7 Fundamental unit in the CAPS format.

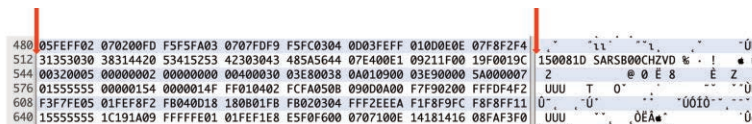


Figura 8 Unità fondamentale nel formato miniSEED.

Figure 8 Fundamental unit in the miniSEED format.

Le parti aggiuntive presenti nel file scritto tramite protocollo CAPS vengono riportate nella Tabella 5.

Campo	Valore in esadecimale	Valore corrispondente in Ascii
File header	48 45 41 44 07 00 00 00 01 00 02 63 6E 74 00	"HEAD cnt" (15 byte)
Unità fondamentale	44 41 54 41 00 02 00 00	"DATA" (8 byte)

Tabella 5 Caratteri aggiuntivi scritti nel file tramite il protocollo CAPS rispetto al file in formato miniSEED standard.

Table 5 Additional parts of the CAPS format compared to the miniSEED format file.

Dall'archive generato tramite protocollo CAPS è possibile quindi ottenere file in formato miniSEED standard semplicemente eliminando i caratteri aggiuntivi indicati in Tabella 1. La Figura 9 mostra un record sismico dove sono sovrapposte le tracce sia in formato miniSEED originale sia quello generato dopo la modifica del file descritto in precedenza.

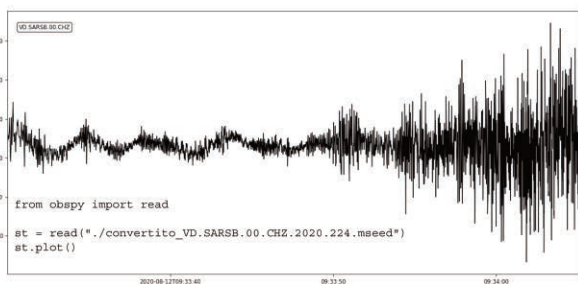


Figura 9 Plot del file scritto tramite protocollo CAPS dopo l'edizione del file eliminando gli elementi presenti in Tabella 5.

Figure 9 Plot of the private CAPS format file of the elements present in Table 5.

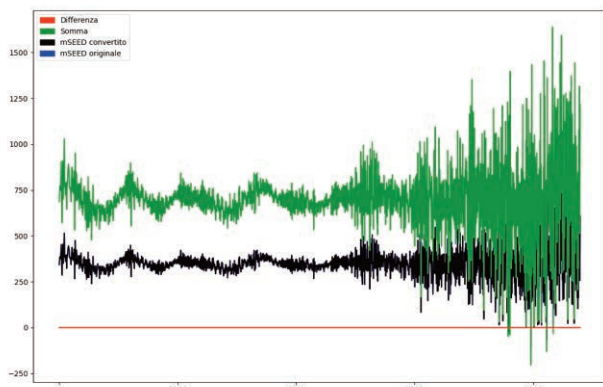


Figura 10 Plot del file scritto tramite protocollo CAPS dopo la modifica del file eliminando gli elementi presenti in Tabella 5 (in nero), del file in formato miniSEED standard (in blue, non visibile), la loro somma (in verde) e la loro differenza (in rosso).

Figure 10 Plot of the private CAPS format file of the elements present in Table 5 (in black), of the file in miniSEED format (in blue, not visible), their sum (in green) and their difference (in red).

Come si può vedere in Figura 10 dal tracciato in rosso e dalla sovrapposizione in nero, il *file* scritto tramite CAPS e modificato come descritto sopra ed il *file* in formato miniSEED *standard* originale combaciano perfettamente. Questa procedura permette di convertire i dati registrati tramite protocollo CAPS al formato miniSEED *standard* a basso costo computazionale. Per questa operazione è stato creato uno *script* per gestire questa conversione quando sia necessaria (p.es., per l'utilizzo di software specifici).

Lo *script* consta di due parti, la prima per convertire l'*header* del *file* in formato CAPS

```
.replace(b"\x48\x45\x41\x44\x07\x00\x00\x00\x01\x00\x02\x63\x6E\x74\x00",b"")
```

e la seconda parte per convertire le unità fondamentali

```
.replace(b"\x44\x41\x54\x41\x00\x02\x00\x00",b"")
```

Di seguito viene riportato un esempio di *script* Python completo che legge il miniSEED e ne esegue la conversione.

```
with open('./rete.stazione.00.EHZ.2020.170.data', 'rb') as f:
    header
    =f.read().replace(b"\x48\x45\x41\x44\x07\x00\x00\x00\x01\x00\x02\x63\x6E\x74\x00",b"").replace(b"\x44\x41\x54\x41\x00\x02\x00\x00",b"")
with open('./conv_rete.stazione.00.EHZ.2020.170.mseed', 'wb') as L:
    data = header
    L.write(data)
```

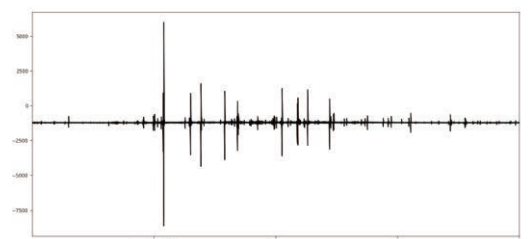
Si osservi che i dati MSEED sono generalmente compressi in modalità Steim1 o Steim2. Per un *byte* di dato compresso tutti i valori possibili sono ugualmente probabili. Ciò significa che una data stringa di 8 *byte* ha una probabilità di occorrenza di $\frac{1}{2^{64}}$. Ovvero circa $\frac{1}{(1,8 * 10^{19})}$. A regime verranno acquisiti *byte* al giorno, quindi la probabilità che si presenti nei dati sismologici di un giorno una specifica stringa di 8 *byte* è dell'ordine di $\frac{1}{10^9}$. Una probabilità sufficientemente bassa. La funzione Python ".substitute" sostituisce tutte le occorrenze di una data stringa, senza alcuna distinzione tra header, parti di controllo o dati effettivi. L'algoritmo di "eliminazione" delle parti non MSEED standard è accettabile perché è bassa la possibilità (ma non totalmente impossibile) che la stringa si presenti nei dati.

Poiché il formato CAPS, ha posizioni fisse per l'*header* e per gli altri caratteri di controllo, una possibilità alternativa che esclude ogni possibilità d'errore sarebbe eliminare i *byte* di troppo che sono in posizione conosciuta.

Abbiamo eseguito un test di conversione su una traccia della durata di 24 ore e la corrispondenza con l'originale in formato MSEED è completa.

Di seguito vengono riportate alcune figure dei *file* convertiti raffrontati con la loro controparte in formato miniSEED *standard*.

Figura 11 File in formato miniSEED *standard*.
Figure 11 Native miniSEED format file.



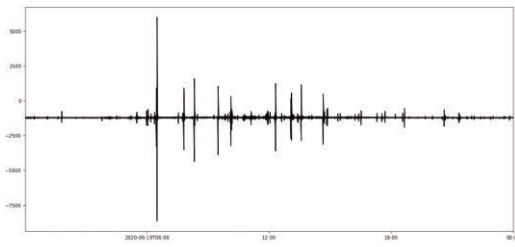


Figura 12 File in formato miniSEED convertito dal file generato tramite protocollo CAPS mediante lo script.
Figure 12 MiniSEED format file converted from CAPS format using the script.

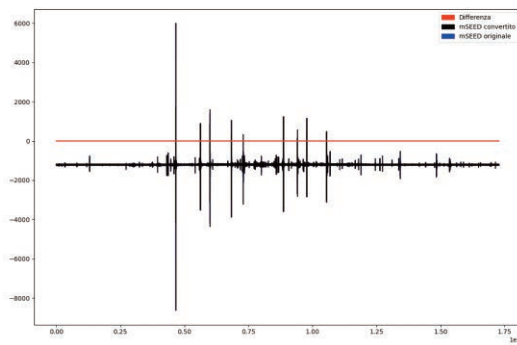


Figura 13 Sovrapposizione e differenza tra file generato tramite protocollo CAPS convertito in miniSEED mediante lo script ed il file in formato miniSEED standard.
Figure 13 Overlap and difference between CAPS format files converted to miniSEED by script and native miniSEED format files.

Si osserva che l'analisi del formato CAPS sopra descritta fa riferimento allo standard precedente l'aggiornamento di formato operato dal produttore in data 17 febbraio 2021 [CAPS nuovo formato, 2021].

2.4 Pipeline processo di analisi

La nostra implementazione di SeisComP prevede l'utilizzo del sistema per l'acquisizione ed elaborazione di dati di diverse reti sismiche locali composte da decine di stazioni localizzate in aree relativamente piccole nel territorio nazionale. Al momento della stesura di questo articolo il sistema acquisisce ed elabora dati provenienti da tre zone di interesse: una localizzata nella pianura Padana nei dintorni della concessione di coltivazione di idrocarburi di Cavone, mentre le altre due si localizzano nella Regione Basilicata nelle concessioni di Val d'Agri e Gorgoglione. La flessibilità di SeisComP ci ha permesso di implementare in parallelo diversi flussi di lavoro per l'elaborazione dei dati (*pipelines*, oppure *processing workflows*). Questa modalità di elaborazione di dati è implementata attraverso l'utilizzo di *aliases* dei moduli *scoutpick* (modulo per il *picking* automatico), *scanloc* (modulo per la localizzazione automatica degli eventi), e *sceval* (modulo per l'identificazione e classificazione degli eventi), come si vede nella Figura 14. Questa possibilità è di estrema importanza a causa della diversità nelle condizioni di operazione che si possono trovare nelle diverse aree di monitoraggio come, ad esempio, il numero di stazioni disponibili ed il livello di rumore.

Ogni *pipeline* consiste in una sequenza di moduli di elaborazione che operano in parallelo usando configurazioni specifiche per il *picking* automatico, per l'associazione di eventi, e per la localizzazione automatica. In questa maniera, è possibile ottimizzare il monitoraggio in funzione delle caratteristiche specifiche di ogni sito, con tutti i dati che confluiscono nello stesso *database*. La Figura 14 mostra un elenco di diversi moduli attivi, in cui sono evidenziate due *pipeline* di *processing* per l'elaborazione dei dati delle tre diverse aree. Una *pipeline* fa riferimento alla concessione di Cavone, mentre l'altra è comune alle concessioni di Val d'Agri e Gorgoglione. Ad esempio, la *pipeline* 1 ha una configurazione in termini di numero di fasi, *score*, filtri, ecc., caratteristici per una rete locale densa (i.e., alto numero di stazioni), mentre la *pipeline* 2 è stata

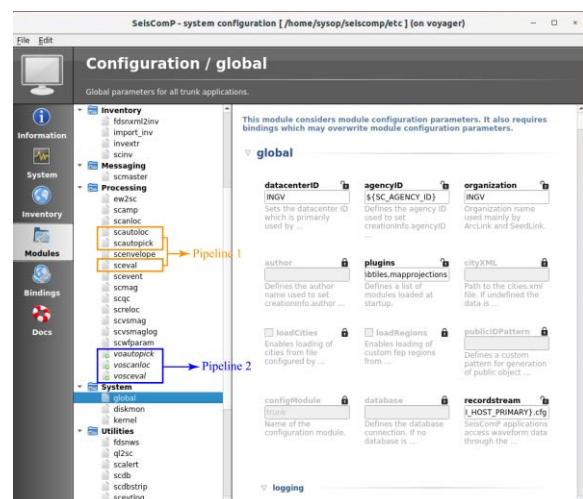
configurata per elaborare i dati di una rete più limitata, caratterizzata da un basso numero di stazioni e filtri adeguati ai segnali con caratteristiche e livelli di rumore diversi da quelli della pipeline 1. La Tabella 6 mostra un confronto di alcuni parametri utilizzati nelle due pipeline implementate.

Filtro	Pipeline 1	Pipeline 2
	Frequenza [Hz] / ordine	Frequenza [Hz] / ordine
Passa alto	3.0 / 3	0.8 / 3
Passa basso	- / -	- / -
STA/LTA		
Lunghezza STA [s]	0.1	0.1
Lunghezza LTA [s]	5	10
Soglia di trigger	3	3
Soglia di de-trigger	1.5	1.5
Tempo pre-event max. ampl. (PEM) [s]	10	10
Dead time [s]	30	30

Tabella 6 Parametri del trigger STA/LTA usati nella configurazione delle due pipeline implementate.
Table 6 STA/LTA trigger parameters used in the configuration of the two implemented pipelines.

Figura 14 Esempio dei moduli in operazione utilizzando due flussi di lavoro paralleli per l'elaborazione automatica dei dati acquisiti.

Figure 14 Example of the modules in operation using two parallel workflows for automatic processing of the acquired data.



Per necessità di monitoraggio è sorta l'esigenza di sviluppare un modulo per il calcolo delle *magnitudo* locali (ML) secondo la formula di Bakun e Joyner [Bakun and Joyner, 1984] e secondo Hutton e Boore [Hutton and Boore, 1987]. Si è considerato come punto di partenza il *plugin* sviluppato in C dal personale dell'Osservatorio del Swiss Seismological Service (SED) per il calcolo della MLh e disponibile presso il loro *repository* GitHub [ML amplitude, 2021]. Il modulo è disponibile come codice sorgente *.cpp*, all'interno del quale si sono apportate le modifiche strettamente legate alle formule della *magnitudo* ed ai parametri A e B da applicare

per le esigenze del nostro monitoraggio. Si riportano come esempi di modifica dei parametri per il calcolo della *magnitudo* secondo Bakun e Joyner [Bakun and Joyner, 1984]:

$$MLbj = \log_{10}(\text{amplitude}) + 1.00 * \log_{10}(\text{hypdistkm}/100) + \text{selected_parameterset.A} * (\text{hypdistkm}-100) + \text{selected_parameterset.B}$$

dove per la *magnitudo* MLbj secondo la formula Bakun e Joyner, vengono scelti i seguenti parametri

$$\text{selected_parameterset.A} = 0.00301 \text{ e } \text{selected_parameterset.B} = 3.0$$

e per la *magnitudo* MLhb secondo la formula di Hutton e Boore, vengono scelti i seguenti parametri

$$\text{selected_parameterset.A} = 0.00189 \text{ e } \text{selected_parameterset.B} = 3.0$$

Dove la base di partenza è fornita dalle formule

$$MLbj = \text{amplitude} + 1.00 * \frac{R}{100} + 0.00301 * (R-100) + 3.0$$

$$MLhb = \text{amplitude} + 1.00 * \frac{R}{100} + 0.00301 * (R-100) + 3.0$$

[Hutton and M. Boore, 1987; Richter, 1935].

Una volta modificato il *file*, esso si compila e si integra in SeisComP (come un *plugin*) per poter essere utilizzato in fase di stima della *magnitudo* dell'evento localizzato.

2.5 Flussi di dati

Ai fini di stimare la capacità di *storage* del sistema di archiviazione e per stimare la massima durata di vita del sistema in funzione del flusso di dati in ricezione sono stati fatti due controlli:

- il *monitor* dell'occupazione di *disk space* giornaliero utilizzato (flusso complessivo dei dati ricevuti);
- il flusso complessivo dei dati in acquisizione (flussi dati provenienti dai singoli canali).

2.5.1 Flusso complessivo dei dati ricevuti

Come primo approccio, poi implementato come controllo standard, viene monitorato lo spazio disco occupato in funzione del tempo. Come valore di riferimento è stato osservato l'ammontare medio dello spazio disco complessivo occupato giornalmente, ottenendo un valore di 5.7 GB per tutte le reti in ricezione al CAD-BO: VA, VO, TP, VD, IV, GE.

2.5.2 Flussi dati provenienti dai singoli canali

Come secondo approccio, calcoliamo prima il flusso teorico dei dati per ogni stazione sismica, in funzione della dinamica e i *settings* dei parametri dell'acquisitore (Tabella 7). Dovuto al fatto

che tutti i datalogger comprimono i dati in compressione Steim1 o Steim2 il valore reale dell'ammontare dei dati si riduce notevolmente (fino ad un fattore 2.5) secondo la rumorosità del sito e del livello di sismicità (Tabella 8).

La quantità dei dati registrati da ogni stazione (*DATA AMOUNT*) dipende da:

- numero di sensori (dipende dalla rete sismica);
- tipo di sensori (p.es. *Short Period, Broad Band* [Seed Manual, 2012]);
- ACQ: tipo di acquisitore (p.es. 24 bit);
- SMP: passo di campionamento (smp) \Rightarrow data streams [Seed Manual, 2012];
- NCHAN: numero di canali (p.es. 3);
- TIME: periodo di acquisizione (secondi).

In Tabella 7 vengono riportati i codici di canale e le bande corrispondenti.

Codice	Banda/tipo strumentazione	periodo (s)	campionamento (Hz)
SH	sismometro a breve periodo	< 10	10 \leq sps < 80
EH	sismometro a breve periodo banda estesa	< 10	80 \leq sps < 250
HH	high-gain sismometro a larga banda	\geq 10	80 \leq sps < 250
HN	high-gain accelerometro	\geq 10	80 \leq sps < 250

Tabella 7 Codici e bande corrispondenti [Seed Manual (2012)].

Table 7 Codes and corresponding bands [Seed Manual (2012)].

$$ACQ \text{ (Byte)} * SMP \text{ (Hz)} * NCHAN * TIME \text{ (s)} = \text{Data Amount}$$

Bit	Byte	Hz/campioni	canali	secondi	data amount (MB/periodo)
24	3	100	1	3600	1.08
24	3	50	3	86400	38.88
24	3	100	3	86400	77.76
24	3	125	3	86400	97.2
24	3	200	3	86400	155.52
24	3	250	3	86400	194.4

Tabella 8 Dettagli del Data Amount.

Table 8 Details of the Data Amount.

Il nostro sistema di acquisizione ora riceve 6 GB di dati di forma d'onda al giorno; questa è la somma di dati acquisiti da stazioni appartenenti a 7 reti sismiche che operano nelle aree di

interesse per il monitoraggio (in totale, 60 stazioni e 282 canali appartenenti alle reti GE, IV, IX, TP, VA, VD, VO).

Per la gestione e pianificazione dell'archiviazione dati progettata nel futuro, sono state fatte delle stime di occupazione dello *storage array*, confrontando il flusso di dati giornaliero teorico con quello reale. La Tabella 8 illustra per ogni singola rete (colonna 1): il tipo di sensore (*velocimetro* o *accelerometro*), il numero di stazioni, il numero di canali e il passo di campionamento (colonne 2-6). La colonna 7 indica il flusso massimo teorico di dati giornaliero per ogni singolo sensore delle stazioni presenti, affiancato dal flusso complessivo per ogni rete sismica. Per i 60 sensori/282 canali delle 7 reti sismiche (inclusa nel calcolo anche la rete VO di Cavone) il flusso dati ammonta circa a 13 GB/giorno. Considerando che la dinamica di acquisizione utilizzata dipende dalla qualità della stazione, dall'attività sismica e dalle fonti di rumore sismico naturale e antropico, il reale flusso medio dei dati registrati è minore: la colonna 9 mostra dei valori medi dell'ammontare reale che complessivamente si attesta attualmente intorno a 5.5 - 6 GB/giorno, oppure 2.2 TB/anno.

Rete	Sensor	n sta	n ch	nsta* nch	smp (Hz)	MaxByte		real	Real Data
						sta (MB/d)	all (MB/d)	factor (%)	sta (MB/d)
VA	vel	17	3	51	200	155.52	2643.8	46.7	1235,00
	acc	14	3	42	200	155.52	2177.3	36.3	791,00
VO	vel	4	3	12	200	155.52	622.1	52.1	324,00
TP	vel	12	3	36	200	155.52	1866.2	42.3	790,00
	acc	12	3	36	200	155.52	1866.2	33.4	623,00
IV	vel	15	3	45	100	77.76	1166.4	55.6	648,00
	acc	6	3	18	100	77.76	466.6	43.7	204,00
	acc	2	3	6	200	155.52	311	48.2	150,00
VD	vel	7	3	21	250	194.4	1360.8	42.2	574,00
IX	vel	2	3	6	125	129.6	259.2	104.2	135,00
	acc	2	3	6	125	129.6	259.2	208.3	270,00
GE	vel	1	3	3	100	77.76	77.8	42.4	33,00
total		60		282			13076.64	44.18	5777,00

Tabella 9 Sunto dei dati sismologici ricevuti.

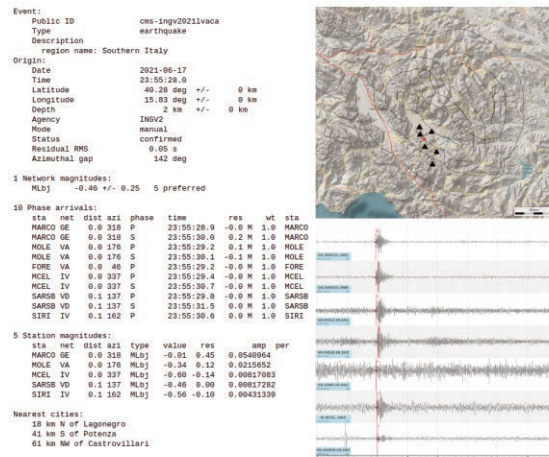
Table 9 Summary of the seismological data received.

3. Messaggistica GDS

Il CMS ha sentito l'esigenza di equipaggiare le macchine di un *tool* per la *reportistica* di eventi rilevati automaticamente o revisionati manualmente da parte del turnista. La scelta è caduta sul *software* Gempa dissemination Server (GDS) [GDS, 2021] fornito da Gempa, integrato con SeisComPro. Questo servizio permette di inviare una *email* con le caratteristiche dell'evento quali *magnitudo*, localizzazione, e fasi che ne hanno permesso la rilevazione. Il *report* viene inviato se l'evento presenta delle caratteristiche di numero di fasi, *magnitudo* e posizione appositamente impostate.

La figura 15 riproduce un esempio di e-mail inviata dal sistema.

Figura 15 Esempio di report via email.
Figure 15 Example of an email report.



4. Conclusioni

L'infrastruttura descritta in questo rapporto tecnico acquisisce ed analizza i dati sismologici sia da reti di pubblico accesso quali la rete sismica nazionale, la rete Geofon, la rete Isnet e la rete Insieme, sia da reti sismiche private. La struttura presenta un alto livello di affidabilità grazie all'infrastruttura gemella che attivamente acquisisce e che, in caso di interruzione di servizio di una sua componente, permette sia di continuare ad acquisire il dato sismologico, che all'operatore di lavorare sull'altra componente. Al fine di prevenire un eventuale malfunzionamento dell'unità di storage, è possibile immaginare una ridondanza della fase di memorizzazione mediante l'introduzione di un ulteriore storage array.

Bibliografia

- Bakun W.H. and Joyner W.B., (1984). *The ML scale in central California*. Bulletin of the Seismological Society of America, 74(5), 1827–1843.
- Braun T., Danesi S. and Morelli A., (2020). *Application of monitoring guidelines to induced seismicity in Italy*. Journal of Seismology, 24, 1015–1028. Doi: 10.1007/s10950-019-09901-7
- Hutton L.K. & M. Boore, (1987). *The ML scale in southern California*. Bull. Seism. Soc. Am., 77, 2074-2094.
- Olivieri M. & Clinton J., (2012). *An Almost Fair Comparison Between Earthworm and SeisComp3*. Seismological Research Letters, 83, 720-727. Doi:10.1785/0220110111
- Rao S., (2014). *Analizzatore software per il formato miniSEED*. Rapporti Tecnici INGV, 285. ISSN 2039-7941.
- Richter C.F. (1935). *An instrumental earthquake magnitude scale*. Bull. Seism. Soc. Am., 25, 1–31.

Sitografia

- Bonding, <https://wiki.debian.org/it/Bonding> data di consultazione: 12/03/2021.
- Braun T., Danesi S., Errico M., Famiani D., Polcari M. and Zerbinato P., (2020). *Progetto di monitoraggio concessione Val d'Agri*. Report 2. http://cms.ingv.it/documenti/Report_2020.1_ValdAgri.pdf data di consultazione: 12/03/2021.

- Braun T., Danesi S. and Famiani D., (2019). *Progetto di monitoraggio concessione Val d'Agri*. Report 1. http://cms.ingv.it/documenti/CMS_REPORT_ValdAgri_2019-01_FINAL.pdf data di consultazione: 12/03/2021.
- CAPS, <https://docs.gempa.de/caps/current/base/introduction.html> data di consultazione 18/06/2021.
- CAPS, <https://gempa.de/products/caps/> data di consultazione: 12/03/2021.
- CAPS nuovo formato, <https://docs.gempa.de/caps/current/base/upgrading.html> data di consultazione: 25/08/2021.
- Capstool, <https://docs.gempa.de/caps/current/apps/capstool.html> data di consultazione: 12/03/2021.
- CMS, <http://cms.ingv.it> data di consultazione: 12/03/2021.
- Common Acquisition Protocol Server - Data Management, <https://docs.gempa.de/caps/4/2020.238/base/archive.html> data di consultazione: 12/3/2021.
- Danesi S., Braun T., Anderlini L., Caciagli M., Errico M., Famiani D., Polcari M. and Zerbinato G., (2020). *Progetto di monitoraggio concessione Val d'Agri*. Report 3. http://cms.ingv.it/documenti/Report_n3_ValdAgri_ufficiale_update.pdf data di consultazione: 12/03/2021.
- Dell Engineering, (2016). *Understanding RAID with Dell SC Series Storage*. Dell EMC. <https://downloads.dell.com/manuals/common/dell-emc-raid-sc-series-storage-2014-cd-ds.pdf> consultato il 17/06/2021.
- Device Mapper Multipathing, https://access.redhat.com/documentation/it-it/red_hat_enterprise_linux/6/html/dm_multipath/mpio_overview data di consultazione: 12/03/2021.
- Earthworm, <https://seiscode.iris.washington.edu/projects/earthworm> data di consultazione: 12/03/2021.
- EarthwormCentral, <http://www.earthwormcentral.org> data di consultazione: 12/03/2021.
- FDSN Web Services, <https://www.fdsn.org/webservices/> data di consultazione: 19/06/2021.
- GDS, <https://www.gempa.de/products/gds/> data di consultazione: 12/03/2021.
- Gitews, <https://www.gitews.org/en/concept/> data di consultazione: 12/03/2021.
- Guida introduttiva a NoMachine, <https://www.nomachine.com/it/guida-introduttiva-a-nomachine> data di consultazione: 12/03/2021.
- Introduction to the Antelope Environmental Monitoring Software. <http://www.crs.inogs.it/antelope/man/html/antelope.1.html> data di consultazione: 12/03/2021.
- ISTI, *Instrumental Software Technologies, Inc.*. <http://www.isti.com> data di consultazione 01/05/2021.
- Libvirt, <https://libvirt.org> data di consultazione: 12/03/2021.
- MiniSEED, <http://ds.iris.edu/ds/nodes/dmc/data/formats/miniseed/> data di consultazione: 12/03/2021.
- ML amplitude, <https://github.com/SeisComp3/seiscomp3/tree/master/src/sed/plugins/magnitudes> data di consultazione: 12/03/2021.
- ObsPy Documentation (1.2.0), <https://docs.obspy.org> data di consultazione: 12/03/2021.
- SAC (Seismic Analysis Code), <https://ds.iris.edu/ds/nodes/dmc/software/downloads/sac> data di consultazione: 12/03/2021.
- Seed Manual, (2012). *Standard for the Exchange of Earthquake Data*. SEED Format Version 2.4, August, 2012. Ahern and Dost (eds.), http://www.fdsn.org/pdf/SEEDManual_V2.4.pdf data di consultazione: 12/03/2021.
- SeedLink, <https://www.seiscomp.de/seiscomp3/doc/applications/seedlink.html> data di consultazione: 12/03/2021.

SeisComP Data Structure (SDS) 1.0, https://geofon-archive.gfz-potsdam.de/software/fdsnws_fetch/SDS.html data di consultazione: 12/03/2021.

Seiscomp, <https://www.seiscomp.de/> data di consultazione: 18/06/2021.

SeisComp3, <https://www.gempa.de/products/seiscompro/> data di consultazione: 19/06/2021.

SeisComPro, <https://www.gempa.de/products/seiscomp3-pro/>, data di consultazione: 12/03/2021.

Understanding Dell SCv2020 RAID and usable spac, <https://www.dell.com/community/Compellent/Understanding-Dell-SCv2020-RAID-and-usable-space/td-p/6042654> data di consultazione: 12/03/2021.

Understanding the Difference Between a Disk, Drive, Volume, Partition, and Image, <https://www.maketecheasier.com/difference-between-disk-drive-volume-partition-image/> data di consultazione: 30/8/2021.

Winston wave server, <https://github.com/usgs/winston> data di consultazione: 19/6/2021.

QUADERNI di GEOFISICA

ISSN 1590-2595

<http://istituto.ingv.it/le-collane-editoriali-ingv/quaderni-di-geofisica.html/>

I QUADERNI DI GEOFISICA (QUAD. GEOFIS.) accolgono lavori, sia in italiano che in inglese, che diano particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari che necessitano di rapida diffusione nella comunità scientifica nazionale ed internazionale. Per questo scopo la pubblicazione on-line è particolarmente utile e fornisce accesso immediato a tutti i possibili utenti. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi. I QUADERNI DI GEOFISICA sono presenti in "Emerging Sources Citation Index" di Clarivate Analytics, e in "Open Access Journals" di Scopus.

QUADERNI DI GEOFISICA (QUAD. GEOFIS.) welcome contributions, in Italian and/or in English, with special emphasis on preliminary elaborations of data, measures, and observations that need rapid and widespread diffusion in the scientific community. The on-line publication is particularly useful for this purpose, and a multidisciplinary Editorial Board with an accurate peer-review process provides the quality standard for the publication of the manuscripts. QUADERNI DI GEOFISICA are present in "Emerging Sources Citation Index" of Clarivate Analytics, and in "Open Access Journals" of Scopus.

RAPPORTI TECNICI INGV

ISSN 2039-7941

<http://istituto.ingv.it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html/>

I RAPPORTI TECNICI INGV (RAPP. TEC. INGV) pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico come manuali, software, applicazioni ed innovazioni di strumentazioni, tecniche di raccolta dati di rilevante interesse tecnico-scientifico. I RAPPORTI TECNICI INGV sono pubblicati esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi.

RAPPORTI TECNICI INGV (RAPP. TEC. INGV) publish technological contributions (in Italian and/or in English) such as manuals, software, applications and implementations of instruments, and techniques of data collection. RAPPORTI TECNICI INGV are published online to guarantee celerity of diffusion and a prompt access to published data. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

MISCELLANEA INGV

ISSN 2039-6651

http://istituto.ingv.it/le-collane-editoriali-ingv/miscellanea-ingv.html

MISCELLANEA INGV (MISC. INGV) favorisce la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV. In particolare, MISCELLANEA INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc. La pubblicazione è esclusivamente on-line, completamente gratuita e garantisce tempi rapidi e grande diffusione sul web. L'Editorial Board INGV, grazie al suo carattere multidisciplinare, assicura i requisiti di qualità per la pubblicazione dei contributi sottomessi.

MISCELLANEA INGV (MISC. INGV) favours the publication of scientific contributions regarding the main activities carried out at INGV. In particular, MISCELLANEA INGV gathers reports of scientific projects, proceedings of meetings, manuals, relevant monographs, collections of articles etc. The journal is published online to guarantee celerity of diffusion on the internet. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

Coordinamento editoriale

Francesca DI STEFANO
Istituto Nazionale di Geofisica e Vulcanologia

Progetto grafico

Barbara ANGIONI
Istituto Nazionale di Geofisica e Vulcanologia

Impaginazione

Barbara ANGIONI
Patrizia PANTANI
Massimiliano CASCONI
Istituto Nazionale di Geofisica e Vulcanologia

©2022

Istituto Nazionale di Geofisica e Vulcanologia
Via di Vigna Murata, 605
00143 Roma
tel. +39 06518601

www.ingv.it



Creative Commons Attribution 4.0 International (CC BY 4.0)



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA



GE.MARCO..HHZ
0.01 / 132.01