

# Back to Las Vegas: il nostro viaggio all'AWS re:Invent 2021!

12 Dicembre 2021 - 10 min. read

[AWS re:Invent](#)

È passata una settimana dal termine del re:Invent 2021, l'evento che io chiamo scherzosamente “la fiera mondiale del Cloud”, in quanto momento certamente più atteso dell'anno per chi, come noi, lavora in questo mondo.

Quest'anno certamente ancora più atteso in quanto ha segnato, per moltissime persone, il ritorno a un grande evento tecnologico in presenza dopo due anni esatti (il mio ultimo evento fu in effetti il [re:Invent 2019](#)).

Non mi dilungherò a raccontare come l'esperienza del ritorno in presenza sia stata piacevole da un punto di vista personale e di come abbia permesso di godere al meglio di ogni singolo spunto offerto dalla conferenza, rispetto alla modalità “solo online” a cui siamo stati costretti negli ultimi 24 mesi - anche perchè direi delle ovvietà. Voglio invece fare i complimenti ad AWS per averci creduto (primo fra i “big” dell'IT) ed aver fatto di tutto per organizzare l'evento in presenza pur con le mille incertezze che fino all'ultimo avrebbero potuto modificare i piani, e con le sfide organizzative e logistiche che ne sono derivate. Certo, sono stati necessari dei compromessi, ad esempio sul numero di persone (quest'anno pare fossimo in 20.000 contro i più di 60.000 del 2019) o sull'organizzazione di alcuni eventi specifici, ma in generale è stata una conferenza davvero ben riuscita se consideriamo i presupposti. Anzi, devo dire che con meno di un terzo dei partecipanti rispetto all'ultima volta, ho trovato l'evento molto più “vivibile”.

Come ho già detto molte volte, l'aspetto più interessante del re:Invent non sono certo i - pur interessantissimi - contenuti presentati nei (ben quattro!) keynote e nelle

centinaia di sessioni: la maggior parte di questi contenuti, infatti, sono disponibili online dopo poche ore sotto forma di decine di **articoli riepilogativi** di tutti gli annunci più importanti, oppure di registrazione di **tutte le sessioni comodamente disponibili su YouTube** (che poi va a finire che, proprio perchè si pensa “tanto le potrò guardare quando voglio, appena avrò tempo”, non le si guardano mai, ma questa è un'altra storia...)

Per questa ragione, come ogni anno, non voglio scrivere l'ennesimo elenco di tutte le news annunciate al re:Invent, ma cercherò di raccontare più che altro le impressioni che la conferenza mi ha lasciato e cosa mi porto a casa come spunti per il 2022.

L'aspetto peculiare del re:Invent è la possibilità di “toccare con mano” un mondo con cui normalmente interagiamo filtrati da una web console o da una CLI, avendo la possibilità di conoscere e parlare direttamente con i responsabili dei vari prodotti e servizi, o con gli speaker delle varie sessioni tecniche.



Quando parlo di “toccare con mano” intendo nel vero senso della parola! Ad esempio quello che si vede in questa foto è uno degli switch custom che AWS utilizza nei suoi data center, e che un membro dello staff AWS mi ha raccontato nei dettagli, sopportando con gentilezza il mio bombardamento di domande. *Custom* perchè, mentre il data plane (la parte più a sinistra con il dissipatore metallico più grande), pur raggiungendo numeri mostruosi - se non ricordo male si parla di più di 400 Terabit/s di banda aggregata! - utilizza componenti standard di mercato, il control plane (la scheda al centro, con il dissipatore più piccolo), quello che di fatto deve gestire tutte le chiamate al servizio VPC e conseguenti configurazioni - per ogni VPC, per ogni account AWS, per gli svariati milioni di clienti! - è un ASIC proprietario a marchio Annapurna Labs, l'azienda specializzata in microelettronica che Amazon ha acquisito nel 2015 per spostare il fulcro dell'innovazione a livello hardware (da qui nasce anche

l'AWS Nitro System). Nessun apparato di rete “commerciale” ha un control plane nemmeno lontanamente così performante, semplicemente perchè on-premise non ce n'è mai stato bisogno! È un'esigenza che è nata con l'evolversi del Cloud e quindi AWS ha fatto prima a “farselo in casa”. Ah, quel singolo switch consuma da solo più di 12kW ed è alimentato direttamente in continua da un backplane proprietario montato direttamente nell'armadio a rack, perchè non era possibile posizionare un trasformatore tanto grande al suo interno. Ecco, questo è il tipo di conversazioni che valgono da sole un viaggio intercontinentale!

In generale l'atmosfera di innovazione che si respira (specie per chi arriva da fuori dagli Stati Uniti) è incredibile, grazie anche al fatto che un evento di questa portata riunisce in un unico luogo fisico le persone più “smart” di una buona parte delle aziende tecnologiche più “cool” al mondo (da HashiCorp a MongoDB, da ARM a CloudFlare, da Slack a Netflix, incluse le “italiane” Kong e Sysdig, tanto per fare qualche nome...), persone con le quali si finisce con grande facilità, tra un evento e l'altro della conferenza, a bere una birra e a discutere di temi tecnici e non. Esperienze come questa portano un valore immenso a chi ha la fortuna di poterle vivere, motivo per il quale ogni anno, come beSharp, partecipiamo al re:Invent con una compagine sempre molto numerosa.



Dal punto di vista degli annunci, a mio parere, è stato un re:Invent un po' sottotono; non saprei dire se sia una questione relativa al particolare periodo storico che stiamo vivendo oppure se il mercato del Cloud (e di AWS nello specifico) sia ormai da considerarsi nella sua fase "matura" e quindi non possiamo più aspettarci lo stesso ritmo di innovazione di qualche anno fa. Fatto sta che non ci sono stati annunci che mi hanno fatto "saltare sulla sedia".

Certamente l'annuncio dalle ricadute più interessanti è quello della modalità serverless e on-demand per quattro dei principali servizi del mondo analytics: **Redshift**, **MSK**, **EMR** e **Kinesis data stream**.



In particolare Redshift è uno dei servizi gestiti più apprezzati di AWS perchè permette di abbattere i costi in maniera consistente rispetto ai data warehouse commerciali (leggi Oracle) ma, così come altri servizi simili, ha storicamente un po' sofferto di essere molto VM-centrico. Sono molto curioso di vedere nel dettaglio il modello di scaling on-demand e ovviamente quello di pricing, presupponendo che anche per questo servizio si possa beneficiare dei progressi architetturali fatti da AWS con Aurora e poi Aurora Serverless relativi alla separazione del layer transazionale da quello di storage. (sperando che, **a distanza di più di un anno dall'annuncio**, Aurora Serverless v2 possa entrare presto in general availability... sui tempi di permanenza in preview dei nuovi servizi annunciati - aumentati molto negli ultimi anni - AWS ha parecchio da lavorare!)

Un annuncio decisamente "scontato" è stato quello di Graviton 3, la CPU con architettura ARM prodotta direttamente da AWS, e delle relative **istanze EC2 C7g**.

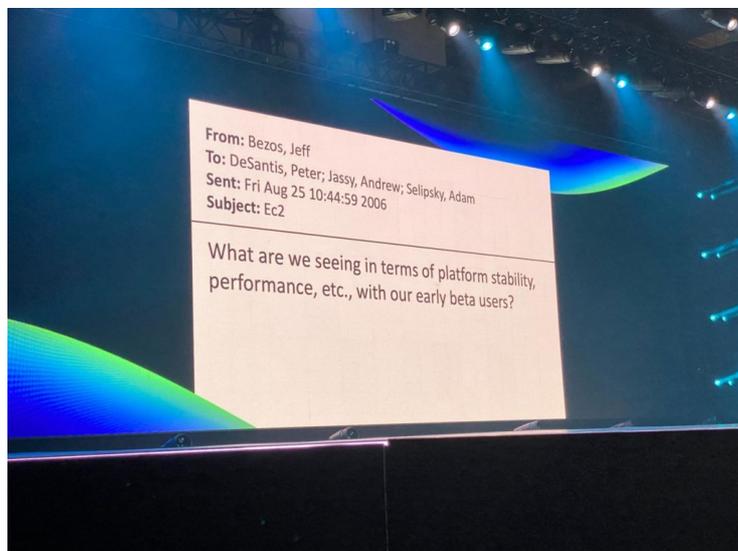


Dico “scontato” non perchè voglia sminuirne il valore (anzi, sulla carta i numeri sembrano particolarmente interessanti, sia dal punto di vista delle performance pure, che dell’efficienza, grazie anche all’utilizzo di memoria DDR5 e del supporto hardware ai calcoli in half-precision per i workload di Machine Learning), ma semplicemente perchè quello delle CPU ARM è uno dei fronti tecnologici al momento più rapidamente in evoluzione e che a ogni iterazione (solitamente annuale) riesce a scaricare a terra aumenti prestazionali e di efficienza a doppia cifra percentuale. E in quest’ottica, infatti, questo annuncio fa coppia con quello delle nuove istanze **EC2 Mac basate sulla CPU M1** di Apple (anche questo largamente atteso). Dopo aver conquistato l’intero comparto mobile in breve tempo grazie all’efficienza energetica, l’architettura ARM negli ultimi 2 anni è entrata di prepotenza nel mondo dei personal computer e dei datacenter (in particolar modo quelli dei Cloud provider) per restarci. “Complice” anche una certa stagnazione evolutiva dell’architettura x86, in particolare sul fronte Intel. AMD invece già da qualche anno con l’architettura ZEN sta continuando a innovare e a tal proposito ho reagito con piacere all’annuncio - finalmente! - dell’**approdo delle CPU EPYC “Milan” (ZEN3) su EC2 (istanze M6a)**. Infatti, se da un lato ARM sta conquistando sempre più fette di mercato in ambito server, molti workload, per varie ragioni sono e saranno ancora per molto tempo vincolati ad architetture x86 (auspicabilmente x86\_64) e, in questo caso, le istanze EC2 con CPU AMD rappresentano a mio parere un’ottima scelta, con una riduzione di costo a parità di prestazioni rispetto alla controparte Intel, e in alcuni casi anche alcune feature in più.

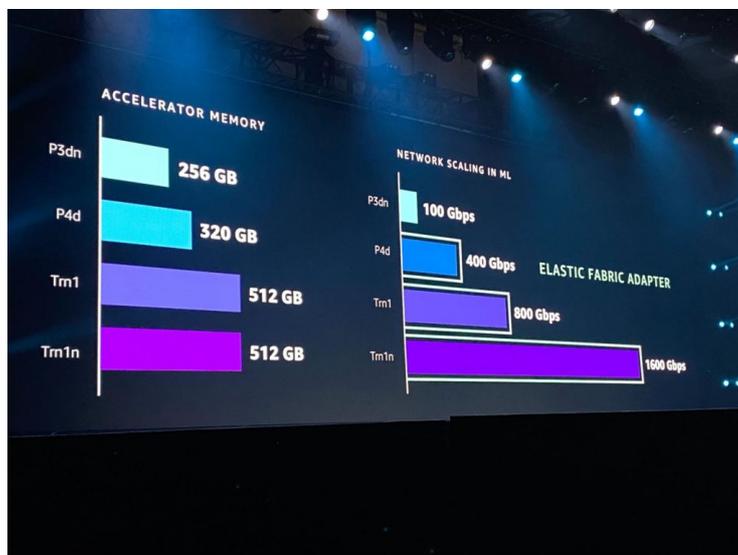
Credo invece che i servizi gestiti di più alto livello basati su EC2 (ad esempio Elasticache o RDS) in futuro saranno sempre più basati su CPU Graviton, dato che in quel caso la CPU sottostante è trasparente all’utente, tranne in casi particolari, come quelli che richiedono l’utilizzo di **RDS Custom**.

Sempre sul fronte infrastruttura e hardware (penso ormai sia chiaro che la parte “under-the-hood” è quella che mi appassiona di più ;-)) da segnalare il rilascio in general availability degli **Outpost in formato “ridotto” a 1 e 2 unità rack** (aka OutPostini) - anch’essi lanciati esattamente un anno fa - ed EBS che guadagna due nuove feature molto interessanti come il “cestino” e **Snapshots Archive**, un nuovo tier di storage a basso costo - atteso a lungo! - per gli snapshot, probabilmente basato su Glacier. E per finire tre chicche dal Keynote di Peter DeSantis:

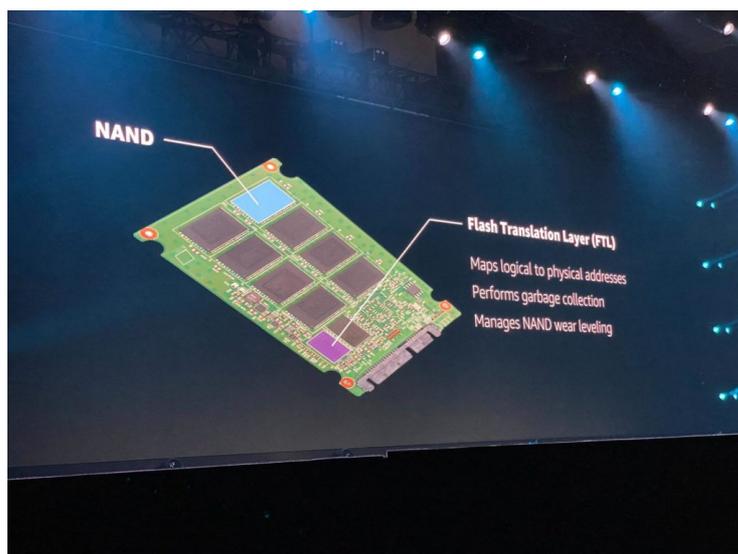
- Lo screenshot (o presunto tale) della mail dell’agosto 2006 con cui Jeff Bezos chiedeva allo stesso DeSantis, Andy Jassy e Adam Selipsky feedback sul beta testing dell’allora neonato servizio EC2. Mostrato in celebrazione del quindicesimo anniversario di EC2. Nota di folklore: il buon Peter è l’unico in questa mail a non essere (ancora?) stato CEO di AWS. Che sia di buon auspicio per lui? :-D



- L’anteprima delle istanze EC2 **Trn1** (ancora in preview) basate sul processore Trainium (silicio proprietario di AWS ottimizzato per il training distribuito di modelli di Machine Learning). I workload per cui questo tipo di istanze è pensato necessitano di una banda molto elevata, ma mai mi sarei aspettato di vedere quanto anticipato per la specifica famiglia Trn1n: 1600 Gbps. UNO VIRGOLA SEI T-E-R-A-B-I-T al secondo su una singola interfaccia di rete di una singola istanza. Pazzesco!



- Una spiegazione molto approfondita di quali sono le sfide per la gestione di storage a stato solido (SSD) in scala “Cloud provider” per garantire prestazioni consistenti e coerenti con una domanda sempre crescente da parte dei clienti. Il “segreto” è nel cosiddetto Flash Translation Layer (FTL), di fatto il firmware del disco. Esattamente come successo per il Nitro System e per gli switch citati poc’anzi, anche qui AWS ha deciso a un certo punto di iniziare a produrre in proprio FTL ottimizzati per la scala, la latenza, le performance e le funzionalità richieste dal Cloud. Nasce così **Nitro SSD**, introdotto **in maniera silente lo scorso anno** e arrivato già alla seconda generazione, che è alla base della **nuova famiglia I4 di istanze EC2** ottimizzate per lo storage ad alte prestazioni.



In effetti, ora che ci penso, un annuncio che mi ha fatto saltare sulla sedia c’è stato, ed è la disponibilità di **OpenZFS come file system per Amazon FSx**, ma mi rendo conto che saremo davvero in pochi nerd al mondo ad emozionarci per una notizia del genere. :-) Sono curioso di testarlo quanto prima e di capire la bontà dell’implementazione come servizio gestito da parte di AWS di quello che resta il miglior file system in circolazione.

Molto interessante anche l'annuncio del Customer Carbon Footprint Tool (coming soon), uno strumento per il cliente per misurare l'impronta ecologica dei propri workload; in piena coerenza con l'annuncio del [sesto pillar dell'AWS Well-Architected Framework: la sostenibilità](#).



In chiusura, ultime menzioni degne di nota:

- [Secrets Detector per CodeGuru Reviewer](#), per non farsi mai più scappare qualche credenziale di troppo in mezzo al codice
- Sul fronte networking evoluto, [AWS Cloud WAN](#), [AWS Direct Connect SiteLink](#) e [AWS Private 5G](#)
- [Control Tower + Terraform](#), vedo positivamente apertura agli standard di mercato
- [AWS Migration Hub](#), per trasformare la tua applicazione in Cloud-native, una API alla volta
- Werner che sfodera [numeri impressionanti a proposito di IAM](#) durante il suo Keynote
- [AWS Mainframe Modernization](#), per non farci mancare proprio nulla

Quello appena concluso è stato il mio decimo re:Invent consecutivo, il nono in presenza a Las Vegas, il primo come [AWS Hero](#) e il quarto come [AWS Ambassador](#) e devo dire che partecipare a questo insieme ai membri di queste due fantastiche community è stata un'esperienza che mi ha davvero arricchito e un privilegio (e non solo per i posti in prima fila riservati durante i Keynote!) Per questo voglio ringraziare in particolar modo [Ross](#), [Farrah](#), [Elaine](#) e [Matthijs](#).



E poi un grazie particolare va anche ai colleghi di beSharp con cui ho condiviso questo viaggio, immortalati in questa foto al booth di AWS, dove abbiamo presentato [Leapp](#), il nostro progetto Open Source! Non ne avete mai sentito parlare? Fatevi un giro sul [repo GitHub](#) del progetto, provatelo e fateci sapere cosa ne pensate!



Vi saluto con questa splendida vista della Strip - quanto mi è mancata Las Vegas! - non senza anticiparvi, per creare un po' di suspense, che uno dei temi principali di cui mi sono interessato quest'anno al re:Invent è stato il [Quantum Computing](#)... ma su questo ne parleremo su queste pagine, prossimamente! Stay Tuned!

#Proud2beCloud



## **Simone Merlini**

CEO e co-fondatore di beSharp, Cloud Ninja ed early adopter di qualsiasi tipo di soluzione \*aaS. Mi divido tra la tastiera del PC e quella a tasti bianchi e neri; sono specializzato nel deploy di cene pantagruliche e nel test di bottiglie d'annata.

---

Copyright © 2011-2021 by beSharp srl - P.IVA IT02415160189