

Gen AI meets Storage: RAG is the next Killer App

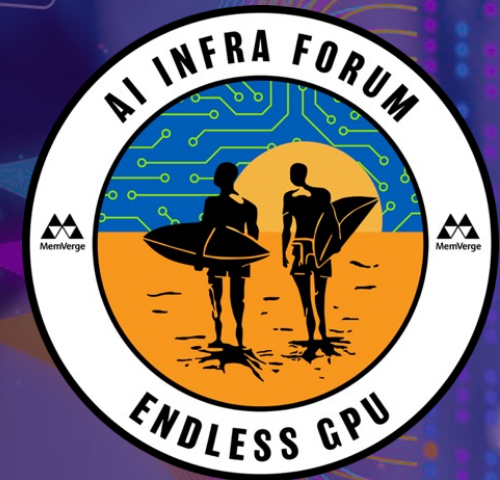
INFINIDAT

Transform private unstructured data into Intelligence with Security, Speed, Scalability



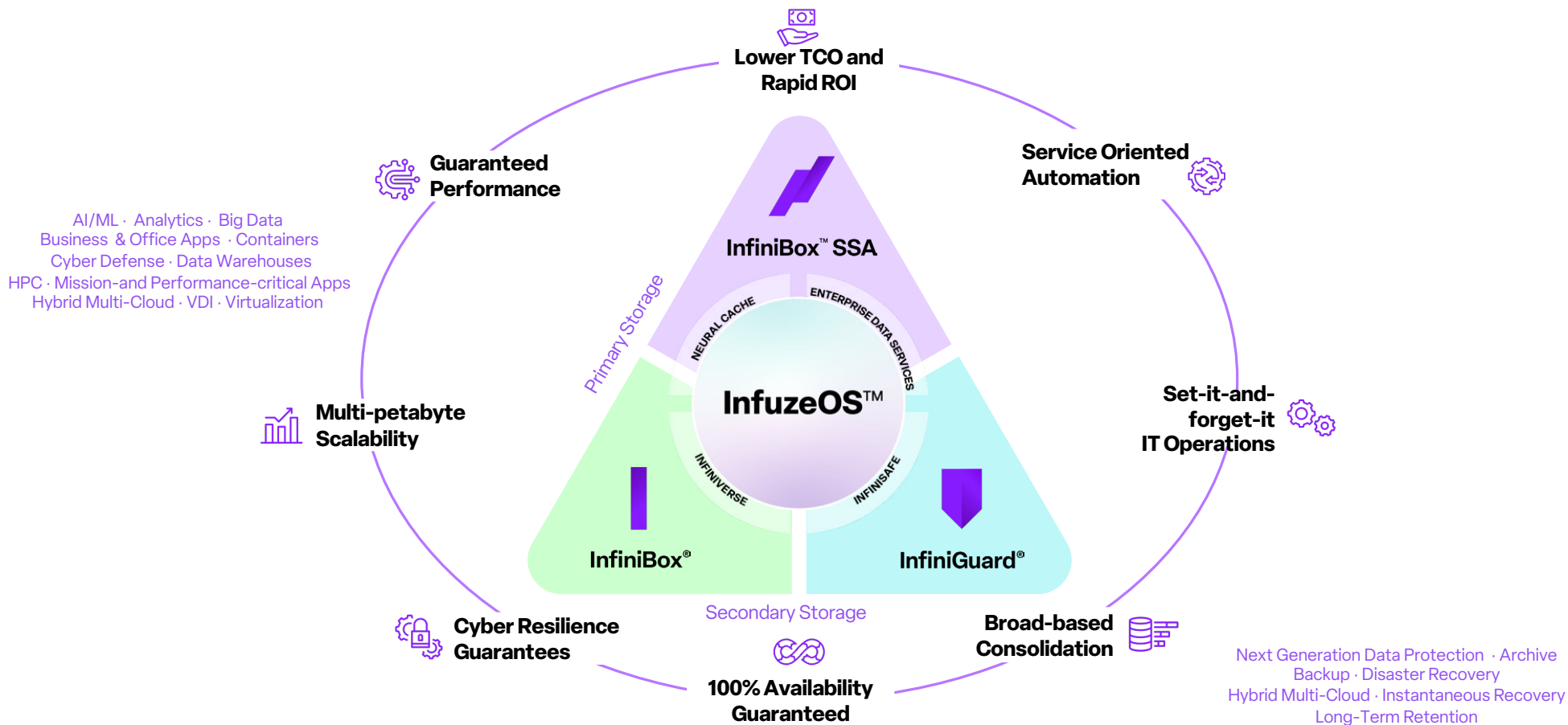
Eric Herzog – CMO

ehertzog@Infinidat.com

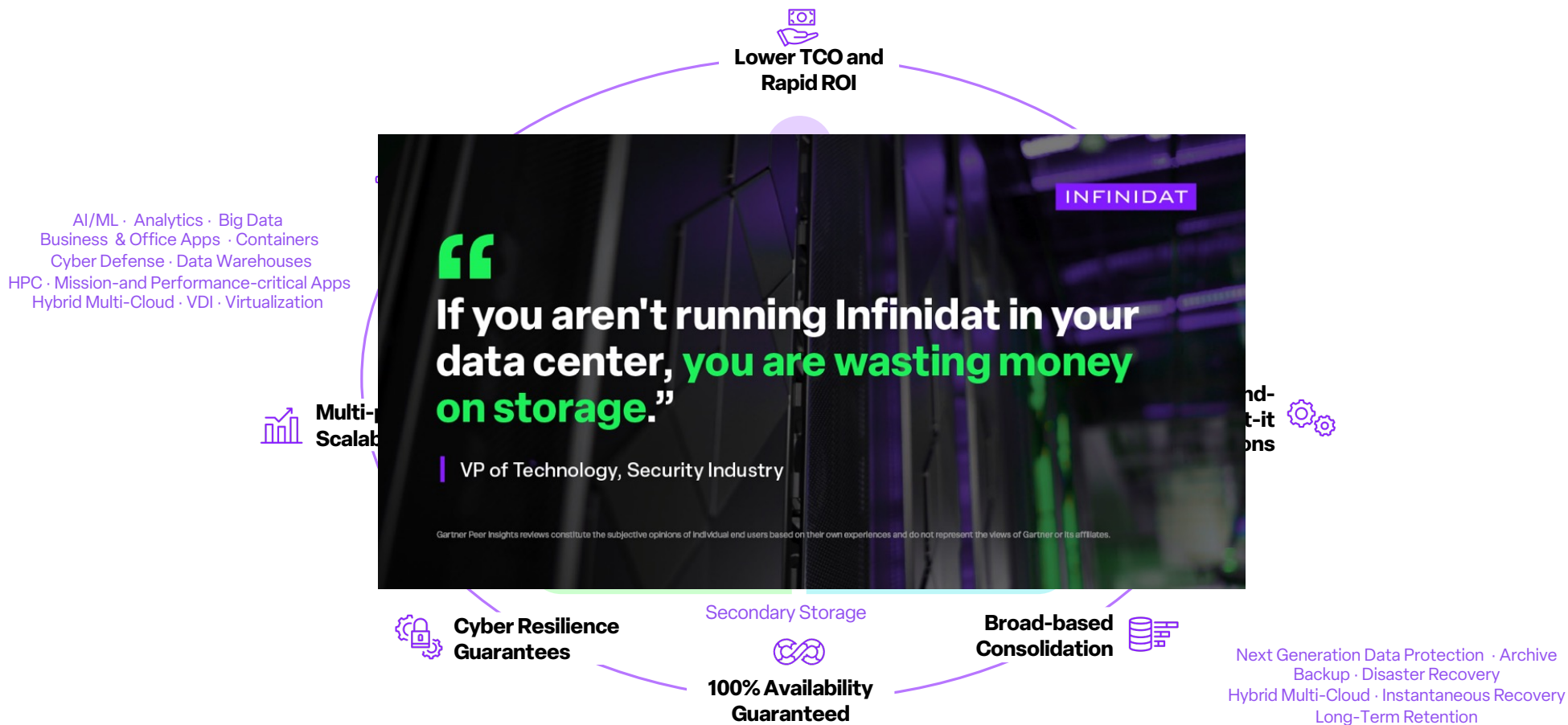


March 2025

Review: InfuzeOS - One SDS Architecture, All Platforms



Review: InfuzeOS - One SDS Architecture, All Platforms



Industry Acclaim: 2024

GIGAOM



InfiniBox® and InfiniBox™ SSA II a Leader and Outperformer - 2024 Radar Report for Primary Storage For Large Enterprises



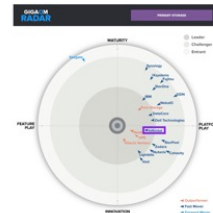
Storage Product of the Year: InfiniBox™ SSA II Gold Medal Award for "The best enterprise storage arrays of 2023"



InfiniBox™ SSA II Named One of Top 5 Cyber Secure High-End All-Flash Arrays for Enterprise Storage



Six Time Gartner® Peer Insights Customers' Choice Award Overall - 4 times North America - 2 times



InfiniBox® and InfiniBox™ SSA a Leader - GigaOm Radar Report for Primary Storage v.5



CRN® Storage Vendor 100 for the Fifth Year in a Row



Capacity Storage Vendor of the Year Winner - Storage Awards - Capacity Storage Vendor of the Year



5-Star Rating in CRN® Partner Program Guide for the 3rd Consecutive Year



Editor's Choice - Product Winner - Editor's Choice - Product - InfiniGuard® Runner-up - Storage Performance, Backup and Optimization



1 of the "Top 10 Leading Enterprise Storage Solution Providers of 2024"



Winner: Channel Storage Company of the Year and Cloud Security Innovation of the Year Runner up: Storage Hardware Innovation of the Year



InfiniBox® and InfiniGuard® Named One of Top 5 2PB+ Cyber Secure Backup Targets in the World



Infinidat Named a 2024 CRN® Tech Innovator Winner for InfuzeOS™ Cloud Edition, and a Finalist for InfiniSafe® Cyber Detection



Eric Herzog - CMO to Watch 2024 by the CMO Alliance



InfiniBox® G4 "Best of Show" for "Most Innovative Hyperscaler Implementation"



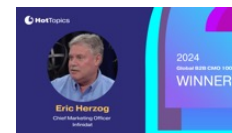
Eric Herzog - 2024 CRN® Inclusive Channel Leaders



Eric Herzog - Top 10 CMO/Marketer of the Year 2024



Infinidat - Finalist for Infrastructure and Storage Manufacturer of the Year



HotTopics - Eric Herzog - Global B2B CMO 100 Award 2024

GIGAOM



InfiniBox® and InfiniBox™ SSA II a Leader and Outperformer - 2024 Radar Report for Primary Storage Midsize Businesses



Steve Sullivan - 2024 CRN® Channel Chief



James 'JT' Lewis - 2024 CRN® Channel Chief - EMEA and APAC



Lynsy Marshall Top 100 Channel Marketing Leader - Channel Insider



Product of the Year Finalist - SSA Express



Infinidat named one of the World's Top 50 Marketing Teams



Lynsy Marshall CRN 2024 Women of the Year Finalist



One of 2024's 50 Coolest Software-Defined Storage Vendors

Industry Acclaim: 2024

CRN CLOUD 100 2025
Outperformer - 2024
Radar Report for Primary Storage For Large Enterprises

THE CHANNEL CO. CRN STORAGE 100 2024
CRN® Storage Vendor 100 for the Fifth Year in a Row

STORAGE AWARDS THE STORIES XXI WINNER
Capacity Storage Vendor of the Year
Winner - Storage Awards - Capacity Storage Vendor of the Year

THE CHANNEL CO. CRN TECH INNOVATORS WINNER 2024
Infinitat Named a 2024 CRN® Tech Innovator Winner for InfuzeOS™ Cloud Edition, and a Finalist for InfiniSafe® Cyber Detection

theCUBE TECHNOLOGY INNOVATION AWARDS 2025
Top Data Storage Innovation FINALIST 2025

2024-25 DCIG TOP 5 GLOBAL EDITION
InfiniBox™ SSA II Named One of Top 5 Secure High-End Enterprise Storage

Gartner Peer Insights Customers' Choice 2024 North America
Six Time Gartner® Insights' Choice Award
InfiniBox® and InfiniBox™ SSA a Leader - GigaOm Radar Report for Primary Storage v.5

STORAGE AWARDS THE STORIES XXI WINNER
InfiniGuard® Named One of Top 5 2PB+ Cyber Secure Backup Targets in 2024

theCUBE TECHNOLOGY INNOVATION AWARDS 2025
Top Data Protection Innovation FINALIST 2025

CRN CLOUD 100 2025
Eric Herzog - CMO to Watch 2024 by the CMO Alliance

FMS Best of Show *****
InfiniBox® G4 "Best of Show" for "Most Innovative Hyperscaler Implementation"

CRN INCLUSIVE CHANNEL LEADERS 2024
Eric Herzog - 2024 CRN® Inclusive Channel Leaders

theCUBE TECHNOLOGY INNOVATION AWARDS 2025
Infinidat - Finalist for Infrastructure and Storage Manufacturer of the Year

CRN CHANNEL CHIEFS REGIONAL 2024
James 'JT' Lewis - 2024 CRN® Channel Chief - EMEA and APAC

100 TOP CHANNEL MARKETING LEADERS channelInsider 2024
Lynsy Marshall Top 100 Channel Marketing Leader - Channel Insider

THE CHANNEL CO. CRN PRODUCTS OF THE YEAR FINALIST 2024
Product of the Year Finalist - SSA Express

ONCON ICON AWARDS TOP 10 MARKETING TEAM AWARDS
Infinitat named one of the World's Top 50 Marketing Teams

THE CHANNEL CO. CRN WOMEN OF THE YEAR FINALIST 2024
Lynsy Marshall CRN 2024 Women of the Year Finalist

CRN CHANNEL CHIEFS 2025
One of 2024's 50 Coolest Software-Defined Storage Vendors

Gen AI - Risk and Reward!

Bad or inaccurate responses leads to:

- Legal and Compliance Issues
- Brand Reputation/Perception Issue
- Customer Satisfaction
- Loss of business to Competitors

Accurate data produces

- Better Customer Experience
- Enhanced Productivity/Efficiency
- Higher Customer Satisfaction
- Innovate and Create Faster
- Cost Reductions and Higher Profits

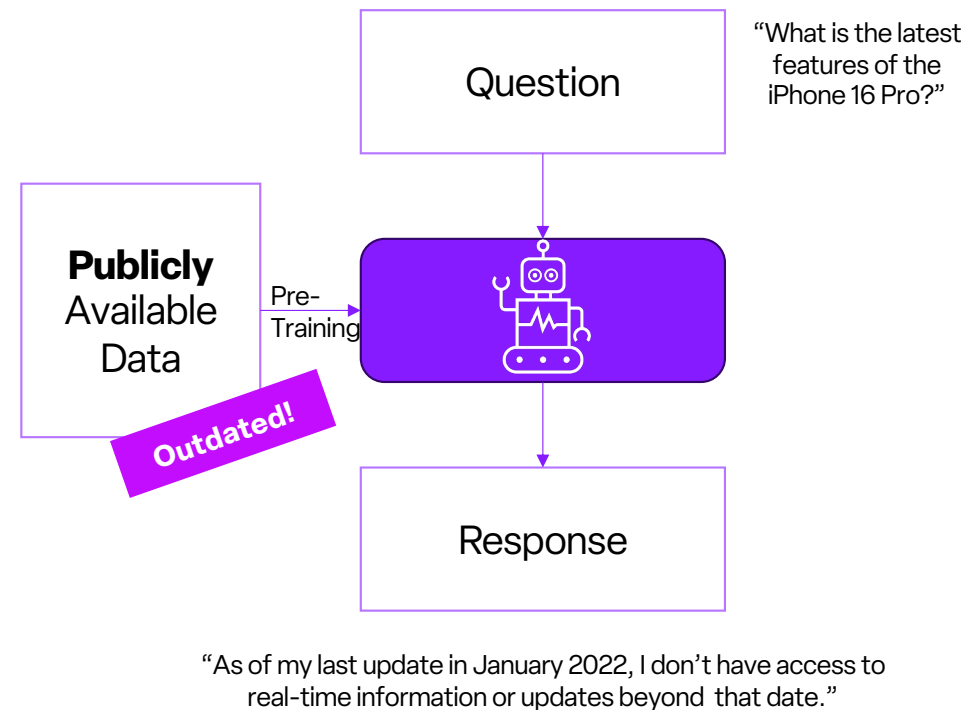
<https://markets.businessinsider.com/news/stocks/amazon-amzn-fails-to-provide-answers-in-ongoing-alexa-ai-delay-1034374823>



“Amazon’s (AMZN) AI upgrade of its Alexa device is expected to be delayed because of one rather inconvenient problem – it gives wrong answers to questions.”

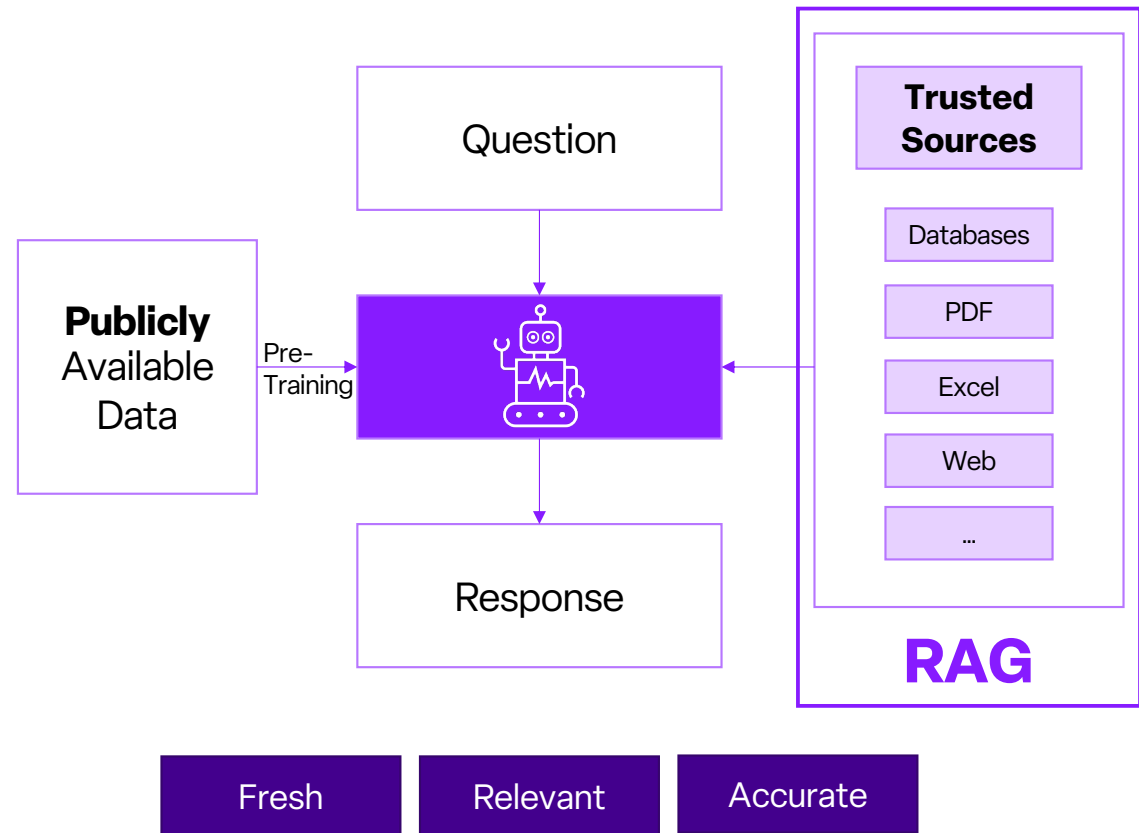
Gen AI – Getting the right answer is a challenge!

- LLMs and SLMs typically trained on publicly available data, lack exposure to an enterprise’s specific and proprietary data
- **Result** - inaccurate and unreliable answers called “hallucinations” appear to be plausible answers but are incorrect or outdated
- Leads to misinformed business decisions, providing customers with wrong information which can damage the enterprise’s reputation
- Gen AI needs continuous augmentation with new, trusted, and accurate data to enhance and maintain its reliability



R(etrieval)A(ugmented)G(eneration) to the Rescue!

- Retrieval-Augmented Generation (RAG) combines the power of GenAI with an enterprise's private data to unlock fresh insights and streamline workflows
- Integrating information from external sources into LLMs and SLMs can achieve more accurate and specific results



Infinidat Gen AI Solution

On-Premise



Replication



InfuzeOS
Cloud Edition

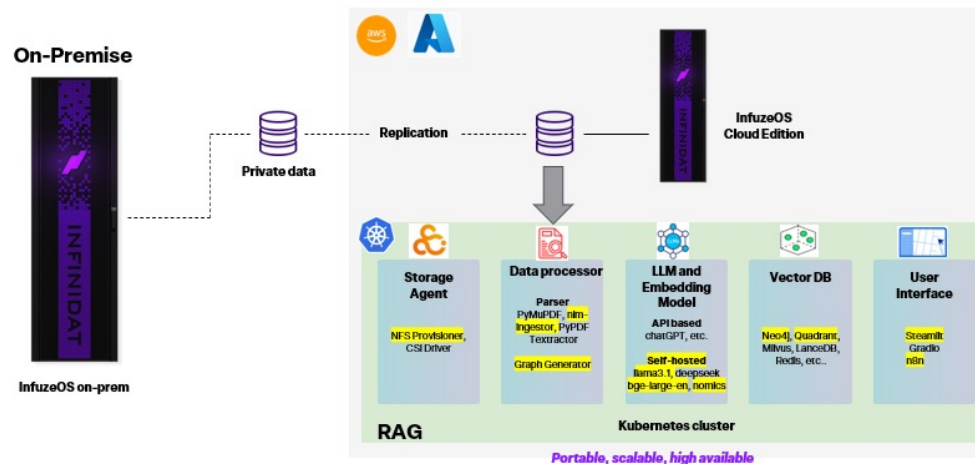
- Adding RAG to benefit Gen AI should work with existing infrastructure in the majority of cases
- Leveraging your existing storage eco-system is key
- Automate and orchestrate the deployment from data center to hybrid multi-cloud
- Easily integrate other storage environments as well (NFS/File)
- Agentic RAG adds new dimensions of flexibility and scalability



Building an Agentic AI Workflow with Infinidat

Day 2 of AI Infra Forum Open-source Projects

Wei Wang - Technology Integration Specialist



Gen AI Meets Storage – More Challenges!

- AI initiatives are hitting a wall at the storage layer in a numerous areas
- Massive amount of checkpoint writes will overrun the storage array
- Expensive GPUs are waiting for data! Bugged Down!
- Huge power usage requirements for AI systems
- LLMs are susceptible to cyber attack

Goldman Sachs | What We Do | Insights | Our Firm | Careers | Investor Relations | Pressroom

ARTIFICIAL INTELLIGENCE

AI is poised to drive 160% increase in data center power demand

NSFOCUS

CLOUD-DELIVERED SERVICES ▾ PRODUCTS ▾ SOLUTIONS ▾ SUPPORT & SERVICES ▾ RESOURCES ▾ NEWS AND EVENTS ▾ COMPANY ▾

Decoding the Double-Edged Sword: The Role of LLM in Cybersecurity

Reuters | World ▾ Business ▾ Markets ▾ Sustainability ▾ Legal ▾ Breakingsviews ▾ Technology ▾ Investiga

Microsoft deal propels Three Mile Island restart, with key permits still needed

By Reuters

September 21, 2024 11:58 AM CDT - Updated 5 months ago

□ Aa ↶



InfiniBox Advantage for AI

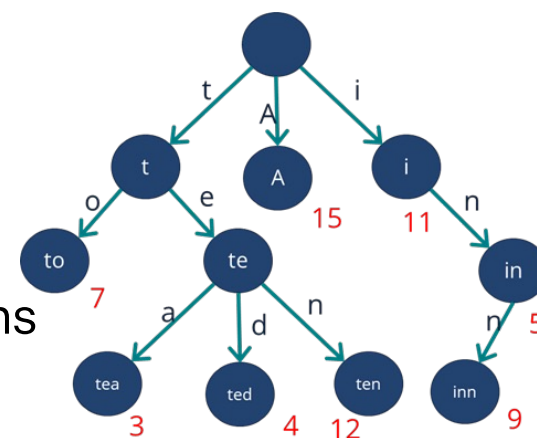
- RAG-ready!
 - Many of our current customers use their InfiniBox in mission-critical applications that utilize Databases which supports storing and retrieving vector data
- Neural Cache
 - Latency impacts AI response to end users and workflows
 - InfiniBox achieves lightning-fast performance with very low latency 35us, will not bog-down the RAG workload process during the Embedding stage



Data Services – Neural Cache



- Infinidat Neural Cache software goes beyond AI/ML to true Deep Learning
- Uses patented Trie technology
- Maintains a history of every write in DRAM
- Analyzes workloads in real time, collects and leverages metadata to improve performance over time
- Stores the data and metadata to create dynamic connections between data, yielding efficiencies in our Deep Learning algorithms for both short- and long-term purposes
- Makes predictions about the future to pre-stage data in cache
- The longer it runs, the faster it gets! (it's learning!)
- **RESULT – Accelerates Application Performance**



InfiniBox Advantage for AI

- N+2 resilience: Active, Active, Active controller technology = 100% availability
- InfuzeOS Cloud Edition provides cloud-based storage operation similar to on-prem enabling storage admins to have the same powerful InfuzeOS in the cloud, to rapidly replicate datasets to AI resources

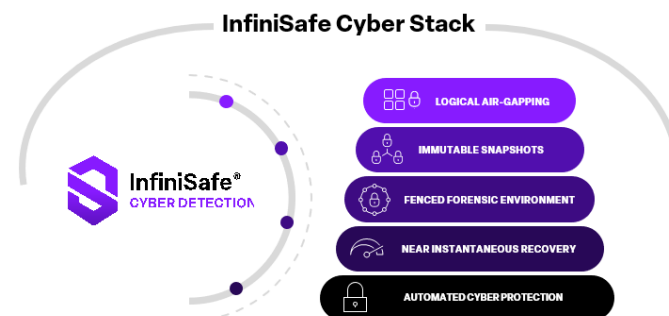


InfiniBox Advantage for AI



InfiniSafe®

- InfiniSafe provides end-to-end cyber storage detection and prevention, with recovery guarantees!
- Secure-by-design approach with three essential InfiniSafe products:
 - InfiniSafe ACP – Automated Cyber Protection – at the “front end” seamlessly and proactively coordinates the detection of abnormal file activity with syslog, SIEM, or SOAR applications
 - InfiniSafe Core Cyber Stack - award-winning InfiniSafe cyber resilience stack
 - Immutable snapshots
 - Local or remote logical air-gaping
 - Fenced forensic environment
 - Near-instantaneous recovery from cyberattacks
 - InfiniSafe Cyber Detection extends cyber prevention further by validating the integrity of immutable snapshots using powerful, ML-based scanning engines



InfiniBox Advantage for AI

- Most power efficient storage in Enterprise-class

| Discover Your OpEx Savings with Infinidat | | |
|---|-------------------------------------|--------------------------------|
| | Enterprise-class Storage Competitor | Infinidat InfiniBox F4400 |
| Max Raw Capacity | 2000 TB | 2000 TB |
| DC Floor Expense | 2 Tiles | 2 Tiles |
| Annual Energy Consumption - kWh (Power and cooling)* | 279,746.22 | 96,360.00 |
| Your Annual Savings with Infinidat | | \$27,507.93 |
| Your Electricity Cost / kWh | \$ 0.15 | 66% More Power Efficient |

66%
MORE
POWER
EFFICIENT
(kWh)

InfiniBox Advantage for AI

- SLA-based Guarantees!



100% Availability Guarantee

Our N+2 controller design and InfiniRAID data protection provide constant monitoring, self-healing, and graceful recoveries from hardware failures on all levels. This design produces a high degree of resilience, enabling Infinidat to guarantee 100% system availability—no downtime!



Performance Guarantee

Infinidat will tailor a set of iron-clad performance service level guarantees based on an understanding of your business use cases and workloads.



InfiniSafe Cyber Storage Guarantees

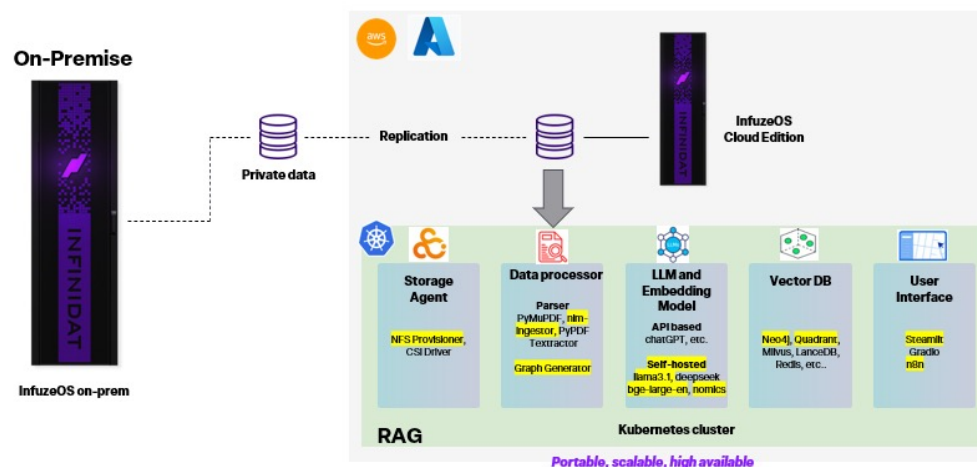
InfiniSafe Cyber Storage Guarantees build on our core InfiniSafe technology and guarantee both recoverability of immutable snapshots and recovery in less than a minute, regardless of data set size.



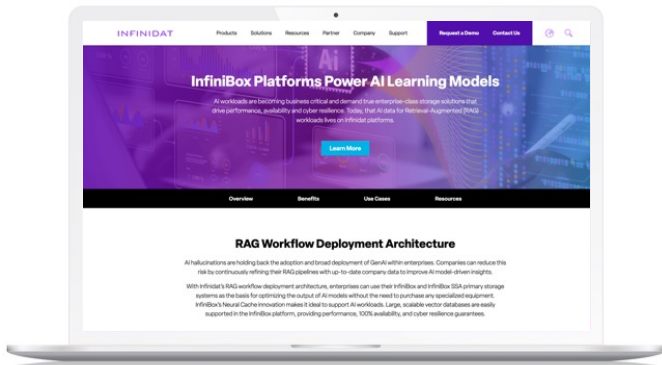
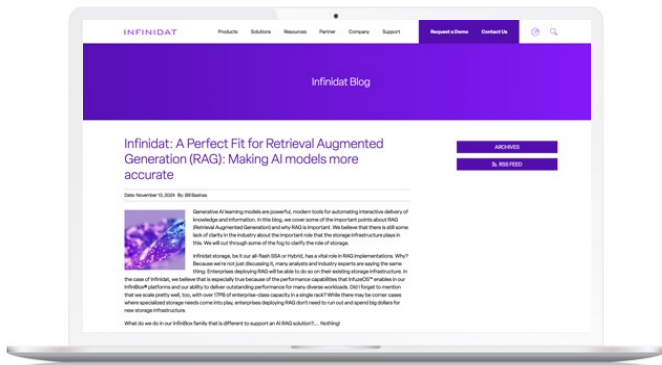
Building an Agentic AI Workflow with Infinidat

Day 2 of AI Infra Forum Open-source Projects

Wei Wang - Technology Integration Specialist



AI Solutions Resources



Blog:

Infinidat: A Perfect Fit for Retrieval Augmented Generation (RAG): Making AI models more accurate

<https://www.infinidat.com/en/blog/infinidat-perfect-fit-retrieval-augmented-generation-rag-making-ai-models-more-accurate>

Analyst Whitepaper:

<https://www.infinidat.com/en/resource-pdfs/infinidat-cures-rag-response-time.pdf>

Analyst Video:

<https://blocksandfiles.com/2025/02/13/how-to-reduce-ai-hallucinations/>

Solution Webpage:

<https://www.infinidat.com/en/artificial-intelligence>



Thank You