
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

CMP312

Run high-performance storage workloads on EC2 storage optimized instances

Amit Shah

Principal Product Manager
AWS

Shruti Gupta

Senior Product Manager
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Storage-driven workloads

Amazon EC2 storage optimized instances

Amazon EBS offerings

Where should I place my workload?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Storage-driven workloads

Applications

Real-time databases
Rideshare
Online gaming
E-commerce and retail AdTech – Segment data and ad bidding

Workload

MySQL
MongoDB
Hadoop
Elastic
Kafka
Oracle
MS SQL server
SAP HANA

Storage

Amazon EC2 instance store
Amazon EBS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS storage portfolio



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 C, M, R, and storage optimized instances

C, M, R instances

C, M, R disk instances

Storage optimized instances

More than

575

Instance types

for virtually every workload and business need

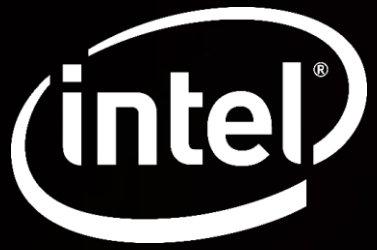
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is a EC2 storage optimized instance?

- Local attached NVMe storage
- Sub-millisecond (<1ms) latency
- High sequential read and write access to very large data sets
- Deliver up to 3.2M random read I/O operations per second (IOPS) to applications

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer choice of processors for storage optimized instances



Intel Xeon Scalable processors

I3	I3en	I4i
Broadwell	Skylake, Cascade Lake	Ice Lake



AWS Graviton processors

Im4gn	Is4gen
AWS Graviton2	AWS Graviton2

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Nitro SSD

HIGH PERFORMANCE, LOW LATENCY SSD CUSTOM DESIGNED BY AWS



Lower latencies

Tightly integrated with the AWS Nitro System to provide 60% lower I/O latency



Improved reliability

Faster firmware updates to improve reliability without any downtime to the instance



Nitro security

All data stored on the disks is encrypted at rest with AES-256 ephemeral keys

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW

Introducing Torn Write Prevention

Improve database performance with Amazon EC2 instance store and Amazon EBS



AWS Nitro SSD

- I4i storage-optimized instance
- Available on Amazon EBS with Nitro-based instances

Database performance

- Up to 30% database performance improvement on transactions per second (TPS)

Cost

- Improved database performance eliminates the need to overprovision or scale up the cluster to meet business growth, reducing TCO

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

I4i: Ice Lake-based storage optimized instance

HIGHEST LOCAL STORAGE PERFORMANCE WITHIN AMAZON EC2



I4i instances

Up to 30 TB of NVMe SSD instance storage with AWS Nitro SSDs

Up to 60% lower I/O latency and up to 75% lower latency variability compared to I3

Up to 75 Gbps NW bandwidth and up to 40Gbps Amazon EBS bandwidth

Up to 55% higher compute performance vs I3

Direct migration from I3

i4i.large		**i4i.32large**
2 vCPU	7 sizes ● ● ●	128 vCPU
16 GiB		1024 GiB
0.468 TB (1x 0.468 TB)		30.0 TB (8x 3.750 TB)

Optimized for workloads such as Oracle and MS SQL databases, NoSQL databases, data indexing and search, and data analytics, with medium size data sets and can benefit from very high compute performance per TB of storage

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Im4gn: AWS Graviton2-based storage optimized instance

BEST PRICE PERFORMANCE FOR STORAGE-INTENSIVE WORKLOADS IN AMAZON EC2



Im4gn instances

Up to 30 TB of NVMe SSD instance storage with AWS Nitro SSDs

Up to 60% lower I/O latency and up to 75% lower latency variability compared to I3

Up to 100 Gbps NW bandwidth and up to 38Gbps Amazon EBS bandwidth

Optimized for workloads that map to 4GB of memory per vCPU

Up to 44% lower price per TB and up to 40% better price performance compared to I3

im4gn.large	6 sizes	im4gn.16xlarge
2 vCPU	● ● ●	64 vCPU
8 GiB		256 GiB
0.937 TB (1x 0.937 TB)		30.0 TB (4x 7.50 TB)

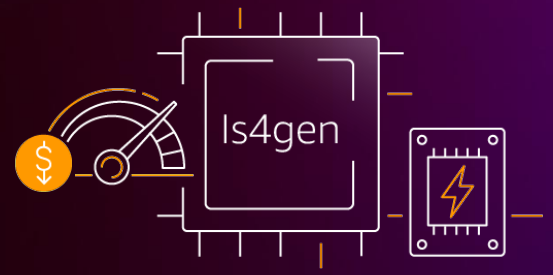
Optimized for workloads such as relational databases, NoSQL databases, data indexing and search, and data analytics, with medium size data sets and can benefit from high compute performance and high network throughput

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Is4gen: Graviton2-based storage optimized instance

THE LOWEST COST PER TB OF SSD STORAGE AND THE HIGHEST DENSITY OF SSD STORAGE PER VCPU



Is4gen instances

Up to 30 TB of NVMe SSD instance storage with AWS Nitro SSDs

Up to 60% lower I/O latency and up to 50% lower latency variability compared to I3en

Up to 50 Gbps NW bandwidth and up to 19Gbps Amazon EBS bandwidth

Optimized for workloads that map to 6GB of memory per vCPU

Up to 15% lower price per TB and up to 48% better compute performance compared to I3en

is4gen.medium	6 sizes	is4gen.8large
1 vCPU	● ● ●	32 vCPU
6 GiB		192 GiB
0.937 TB (1x 0.937 TB)		30.0 TB (4x 7.50 TB)

Optimized for workloads such as NoSQL databases, data streaming, distributed filesystems, and data indexing and search, with large size data sets benefiting from highest storage density per vCPU

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

I-family offerings

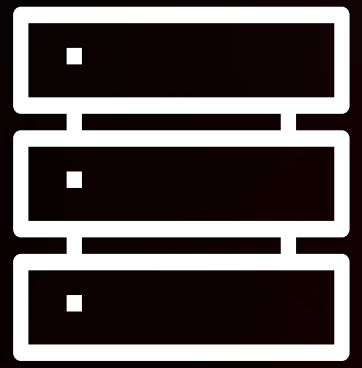
	I4i	Im4gn	I3en	Is4gen
Processor	Ice Lake	AWS Graviton2	Cascade Lake	AWS Graviton2
vCPU:Mem (GB)	1:8	1:4	1:8	1:6
vCPU:Disk (GiB)	1:234	1:468	1:625	1:937
Mem:Disk (GiB)	1:29	1:117	1:78	1:156
$/TB (OD)	$0.366	$0.194	$0.181	$0.154
Optimized for	Best local storage performance per vCPU, specifically the highest vCPU per GB of disk	Highest compute performance	Highest NVMe storage capacity	Lowest $/TB and highest storage/vCPU
Target workloads	Small to medium data sets NoSQL databases (for example, Redis, MongoDB, Aerospike, Couchbase), transactional DBs (for example, MySQL, Oracle DB, and Microsoft SQL Server), Data analytics, Search	Medium data sets, NoSQL databases (KeyDB, ScyllaDB, and Cassandra), RDBMS (MySQL, MariaDB, PostgreSQL), data analytics, and search engines	Large data sets, NoSQL databases (for example, Cassandra, MongoDB, Aerospike), in-memory databases (for example, Redis), scale-out transactional databases, distributed file systems, DWH, indexing, and analytics workloads	Large data sets, NoSQL databases (KeyDB, MongoDB, ScyllaDB, and Cassandra), search engines (Splunk and Elasticsearch), data streaming, and large distributed file systems

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

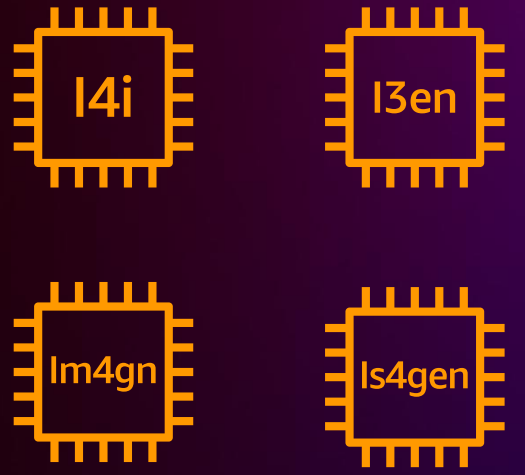
Identify the right instance for your workload



Benchmarking/performance



Memory, disk, and CPU ratios



Instance types and cluster sizing

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Mapping scenarios

- Scale up (1:1)
  - Higher compute performance
  - All the advantages of Graviton2
  - Benefit from new AWS Nitro SSD
- Scale down (*n*:1 at node level)
  - Pick the next instance size down
  - Consider memory constraints
  - Use when optimizing cost
- Scale in (*n*:1 at cluster level)
  - Reduce nodes in cluster
  - Consider cluster quorum
  - Use when optimizing cost
- Any combination of the above
  - Combine scale down/scale in
  - Test different load scenarios
  - Combine to increase cost savings

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer quotes



“We are thrilled with the performance we are seeing from the Amazon EC2 I4i instances which use the new low latency AWS Nitro SSDs. Our testing shows I4i instances delivering an astonishing 2.9x higher query throughput than the previous generation I3 instances. We have also tested with various read and write mixes, and observed consistent and linearly scaling performance.”

Yiftach Shoolman, Co-Founder and CTO – Redis



As the only unified customer experience management platform (Unified-CXM), Sprinklr helps the world's largest and most valuable companies make their customers happier across 30+ digital channels — using industry-leading AI to create insight-driven strategies and better customer experiences.

"We benchmarked our Java-based search workloads on Amazon EC2 Im4gn/Is4gen instances powered by AWS Graviton2. Smaller Is4gen instances offer similar performance compared to larger I3en instances, presenting an opportunity to meaningfully reduce the TCO. We also saw a significant 50% reduction in latency for queries when moving our workloads from I3 to Im4gn instances, indicating a significant 40% price performance benefit. Moving to Graviton2-based instances was easy, taking 2 weeks to complete benchmarking. We are very happy with our experience and look forward to running these workloads in production on Im4gn and Is4gen instances."

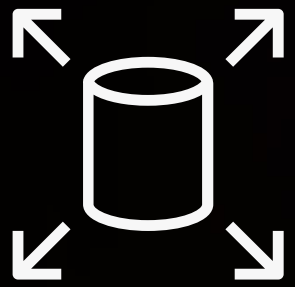
Abhay Bansal, Vice President of Engineering – Sprinklr

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introduction to Amazon EBS

Amazon Elastic Block Store (Amazon EBS) is an easy to use, secure, high-performance, block-storage service designed for use with Amazon EC2, for both throughput and transaction intensive workloads at any scale.



Amazon EBS Volumes

gp2, gp3, io1, io2, io2 Block Express, st1, sc1



Amazon EBS Snapshots

Incremental point-in-time copies of EBS volumes



Data Services

Elastic Volumes, Fast Snapshot Restore, Data Lifecycle Management

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Select volume that best fits your workload

	SSD			HDD	
	gp3 General Purpose	Io2 Provisioned IOPS	io2 Block Express Provisioned IOPS	st1 Throughput Optimized	sc1 Cold Storage
Use-cases	Relational and non-relational databases, containers, big data, streaming, file system, media workflows	Large database workloads and business applications requiring sustained high performance	Performance-sensitive, mission-critical applications and databases requiring sub-millisecond latencies	Big data workloads, data warehouses, log processing, streaming workloads	Large volumes of infrequently accessed data, cost-sensitive workloads
Volume Size	1 GB – 16 TB	4 GB – 16 TB	4 GB – 64 TB	125 GB – 16 TB	125 GB – 16 TB
Max IOPS/ volume	16,000	64,000	256,000	500	250
Max Throughput / Volume	1,000 MB/s	1,000 MB/s	4,000 MB/s	500 MB/s	250 MB/s
Durability	99.8-99.9%	99.999%	99.999%	99.8-99.9%	99.8-99.9%

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EBS snapshots: More than just backup

Initial use case

Backup and recovery for Amazon EBS volumes



Expanded use cases

Share Amazon Machine Images (AMIs) to update instances/OS across accounts and Regions



Disaster recovery and ransomware protection across accounts and Regions



Data Lifecycle Manager – Automation across use cases

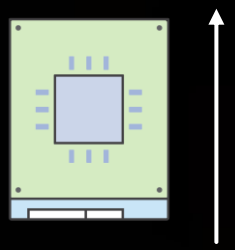
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

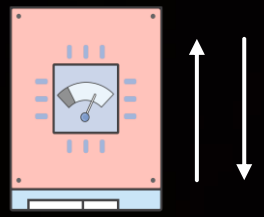
Modify volumes and Fast Snapshot Restore

Elastic Volumes (EV)

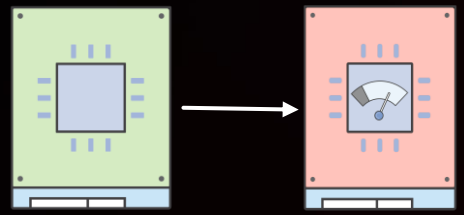
Online volume modification



Increase volume size as application scales



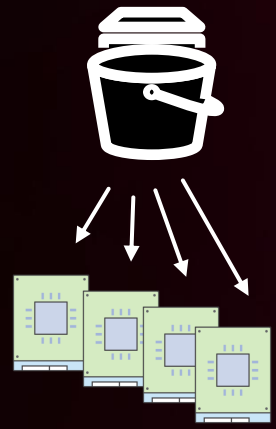
Increase or decreases volume performance to handle changing workload demand



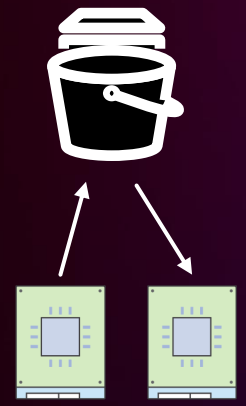
Change volume type to best fit your evolving application needs

Fast Snapshot Restore (FSR)

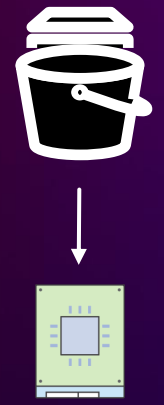
Instantly create fully performant volumes from snapshot



Faster instance launches



Faster and predictable volume copy workflow



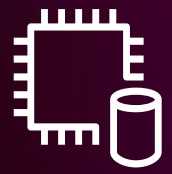
Faster and predictable volume recovery

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

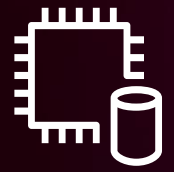
Choose between Amazon EC2 instance store and Amazon EBS volumes

Ultra high performance	Latency < sub-millisecond
	Throughput > 10 GB/second
	IOPS > 260,000

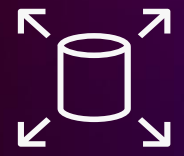


Amazon EC2 instance store

Uptime and data protection	Durability at application level or storage level
	Backup using Amazon EBS Snapshot or Amazon S3



Amazon EC2 instance store



Amazon EBS volumes

Flexibility and features	Storage independent of instance lifecycle
	Scale compute and storage independently
	Features such as EV, DLM, FSR
	Multi-attach multiple instances to one volume

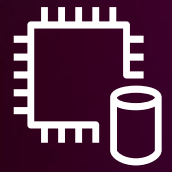
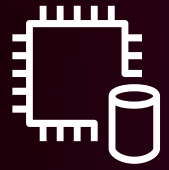
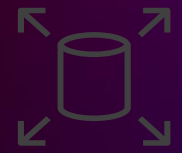


Amazon EBS volumes

aws

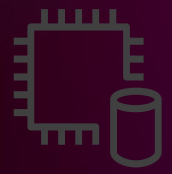
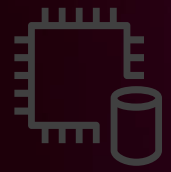
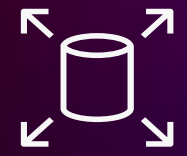
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Example: Aerospike using Amazon EC2 instance store

Ultra high performance	Latency: microseconds Throughput > 10 GB/second IOPS > 260,000	 Amazon EC2 instance store
Uptime and data protection	Durability at application level or storage level Backup using Amazon EBS Snapshot or Amazon S3	 Amazon EC2 instance store  Amazon EBS volumes
Flexibility and features	Storage independent of instance lifecycle Scale compute and storage independently Features such as EV, DLM, FSR Multi-attach multiple instances to one volume	 Amazon EBS volumes

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Example: SAP HANA using Amazon EBS

Ultra high performance	Latency < sub-millisecond Throughput > 10 GB/second IOPS > 260,000	 Amazon EC2 instance store
Uptime and data protection	Durability at application level or storage level Backup using Amazon EBS Snapshot or Amazon S3	 Amazon EC2 instance store  Amazon EBS volumes
Flexibility and features	Storage independent of instance lifecycle Scale compute and storage independently Features such as EV, DLM, FSR Multi-attach multiple instances to one volume	 Amazon EBS volumes

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.