AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

STG201

Amazon EBS under the hood: A tech deep dive

Marc Olson
Senior Principal Engineer
AWS

Prarthana Karmakar
Software Development Manager
AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Storage Portfolio

OBJECT	BLOCK	FILE			BACKUP	
						
Amazon S3	Amazon EBS	Amazon EFS	Amazon FSx for Windows File Server	Amazon FSx for Lustre	Amazon EBS snapshots	AWS Backup

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is Amazon EBS?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

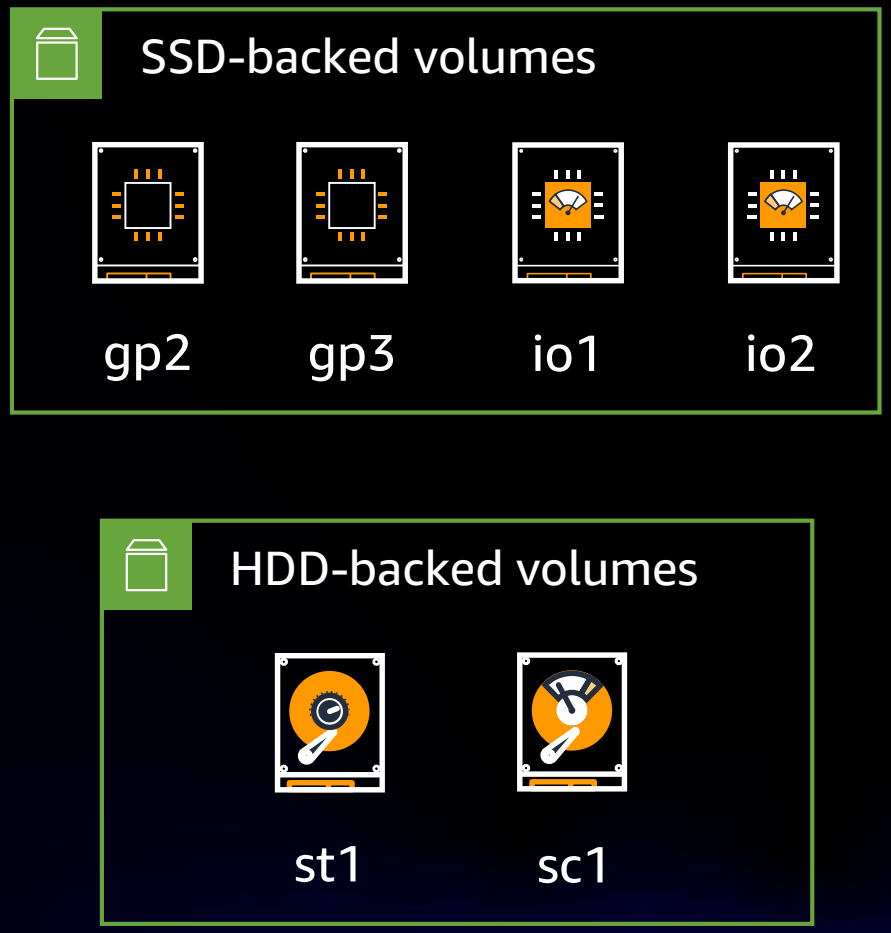
Introduction to Amazon Elastic Block Store (EBS)

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

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

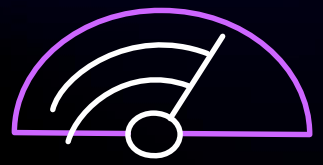
What is Amazon Elastic Block Store (EBS)?

- Block storage **volumes** as a service attached to Amazon EC2 instances
- **Flexible** storage and performance for dynamic workloads such as stateful containers
- Create, attach, and manage volumes through **API**, **SDK**, or **AWS console**
- Point-in-time **snapshots** and tools to automate backup and retention via policies



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why do customers use Amazon EBS?



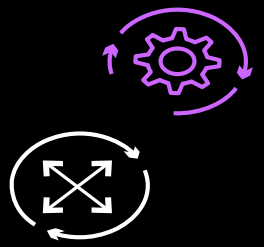
High performance

Up to 64,000 IOPS and 1,000 MiB/s throughput with single-digit millisecond latency



Reliability

99.999% availability and between 0.001% and 0.02% AFR



Scalability

Use a single gigabyte or scale up to petabytes of data for dynamic workloads such as stateful containers



Easy to use

Easily increase capacity, performance, switch volume types, and manage backups



Security

Encrypt all volumes using default encryption keys or custom managed keys



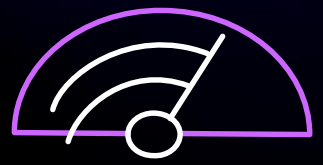
Cost-effective

Pay as low as $0.025/GB-month for HDD and $0.08/GB-month for SSD

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why do customers use Amazon EBS?



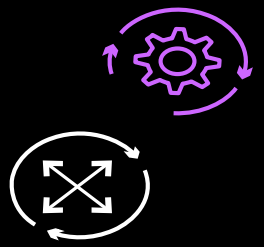
High performance

Up to 64,000 IOPS and 1,000 MiB/s throughput with single-digit millisecond latency



Reliability

99.999% availability and between 0.001% and 0.02% AFR



Scalability

Use a single gigabyte or scale up to petabytes of data for dynamic workloads such as stateful containers



Easy to use

Easily increase capacity, performance, switch volume types, and manage backups



Security

Encrypt all volumes using default encryption keys or custom managed keys



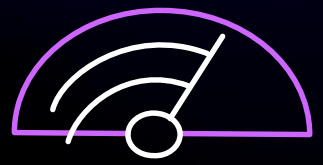
Cost-effective

Pay as low as $0.025/GB-month for HDD and $0.08/GB-month for SSD

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why do customers use Amazon EBS?



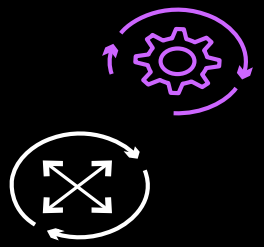
High performance

Up to 64,000 IOPS and 1,000 MiB/s throughput with single-digit millisecond latency



Reliability

99.999% availability and between 0.001% and 0.02% AFR



Scalability

Use a single gigabyte or scale up to petabytes of data for dynamic workloads such as stateful containers



Easy to use

Easily increase capacity, performance, switch volume types, and manage backups



Security

Encrypt all volumes using default encryption keys or custom managed keys



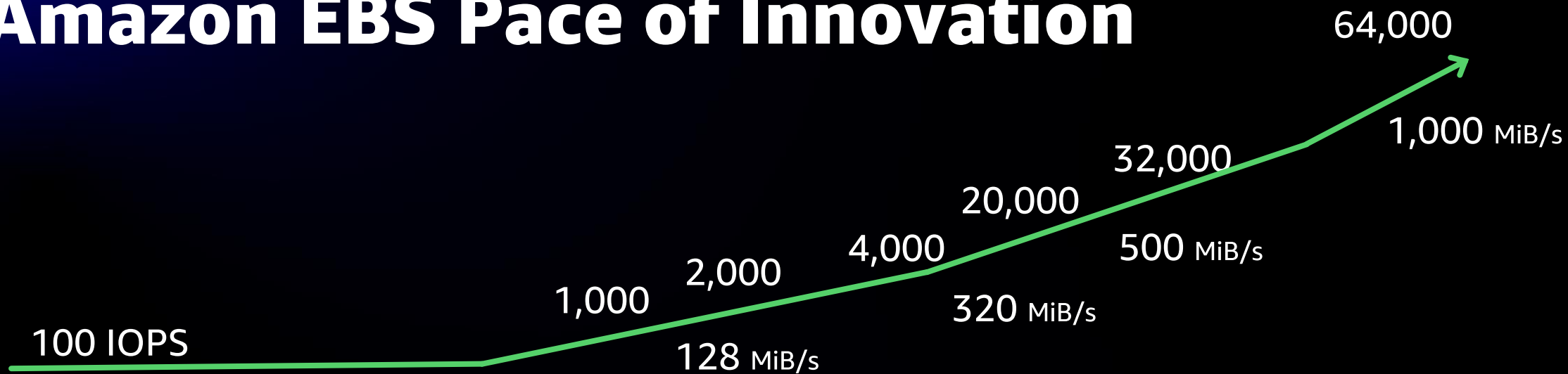
Cost-effective

Pay as low as $0.025/GB-month for HDD and $0.08/GB-month for SSD

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EBS Pace of Innovation



2008 2010 2012 2014 2016 2018 2020

Standard and snapshots
2008

io1
2012

gp2
2014

sc1/st1
2016

Data Lifecycle Manager
2018

Fast snapshot restore
2019

io2
2020

gp3
2020

Amazon EBS optimized
2012

Encrypted volumes
2015

Elastic volumes
2017

Direct APIs
2019

Multi-Attach
2020

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Understand your mission



Databases

PostgreSQL, MySQL

Cassandra, MongoDB



Data & Analytics

Kafka, Splunk, Hadoop,

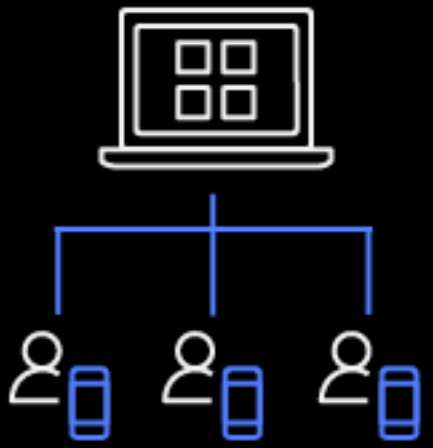
Data Warehousing



Media

Transcoding, Encoding,

Render Farms



File

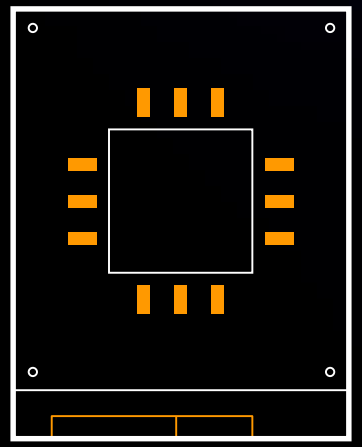
CIFS, NFS

Archive

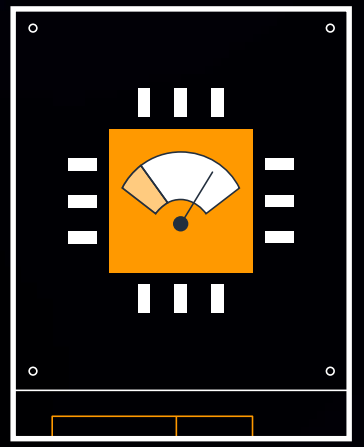
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Select the right volume for your workload

SSD

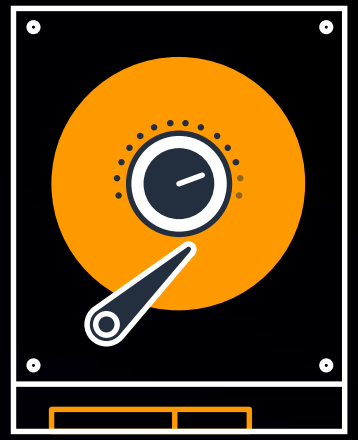


gp3
General Purpose SSD

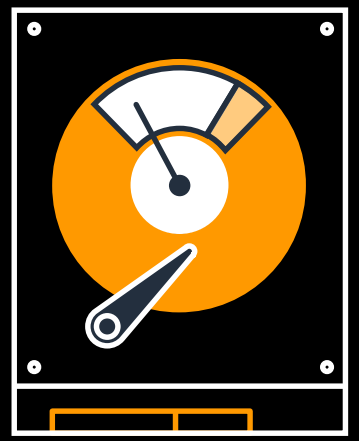


io2
Block Express

HDD



st1
Throughput Optimized HDD



sc1
Cold HDD

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Select the right volume



Databases

PostgreSQL, MySQL

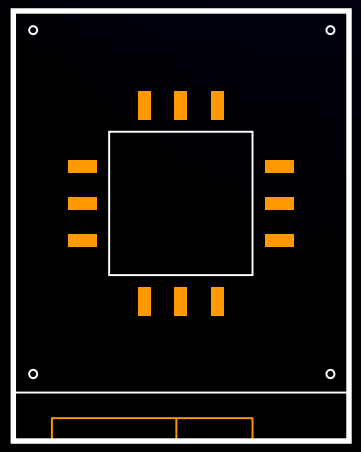
Cassandra, MongoDB

- Typically high performance requirements
- Mostly random I/O
- Journal is sequential
- Highly workload dependent

Recommendation: SSD (+ HDD?)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS volume: General Purpose SSD



gp3

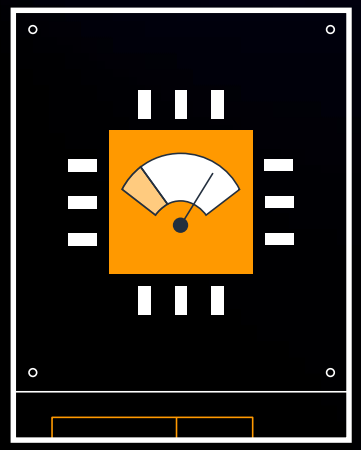
General Purpose SSD

- IOPS: 3,000 – 16,000 IOPS
- Throughput: 128 – 1,000 MiB/s
- Latency: Single-digit ms
- Capacity: 1 GiB to 16 TiB
- I/O Size: Up to 256 KiB (logical merge)

Great for boot volumes, low-latency applications, and bursty databases

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS Volume: io2 Block Express



io2

Block Express

- **Next-generation** architecture
- **4x** throughput and **4x** IOPS
  - Up to 4,000 MiB/s
  - Up to 256,000 IOPS
  - 1,000:1 IOPS to GB
- **4x** volume size up to 64 TB per volume
- **<1 millisecond** latency

Ideal for critical applications and databases with sustained IOPS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“Would my customers or my business be impacted by degradation or an outage?”

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Understand your mission

Tier 1 (critical)

Everything Else

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Select the right volume



Media

Transcoding, Encoding,

Render Farms

- Typically high throughput requirements
- Mostly sequential I/O
- Sustained I/O

Recommendation: Throughput Optimized HDD (st1)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EBS volume: Throughput optimized



st1

Throughput Optimized HDD

- Baseline: 40 MiB/s per TiB, up to 500 MiB/s
- Burst: 250 MiB/s per TiB, up to 500 MiB/s
- Capacity: 125 GiB to 16 TiB
- I/O Size: Up to 1 MiB (logical merge)

Ideal for large-block, high-throughput sequential workloads

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Select the right volume



Data & Analytics

Kafka, Splunk, Hadoop,

Data Warehousing

- Typically high throughput requirements
- Mostly sequential I/O
- Daily periodicity

Recommendation: HDD

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EBS volume: Cold HDD



sc1

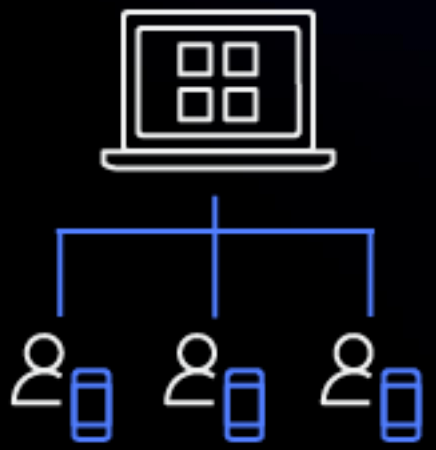
Cold HDD

- Baseline: 12 MiB/s per TiB, up to 192 MiB/s
- Burst: 80 MiB/s per TiB, up to 250 MiB/s
- Capacity: 125 GiB to 16 TiB
- I/O Size: Up to 1 MiB (logical merge)

Ideal for sequential throughput workloads, such as logging and backup

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Select the right volume



File

CIFS, NFS

Archive

- Typically low throughput requirements
- Bursty workloads
- Cost-sensitive

Recommendation: Cold HDD (sc1)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Select the right volumes

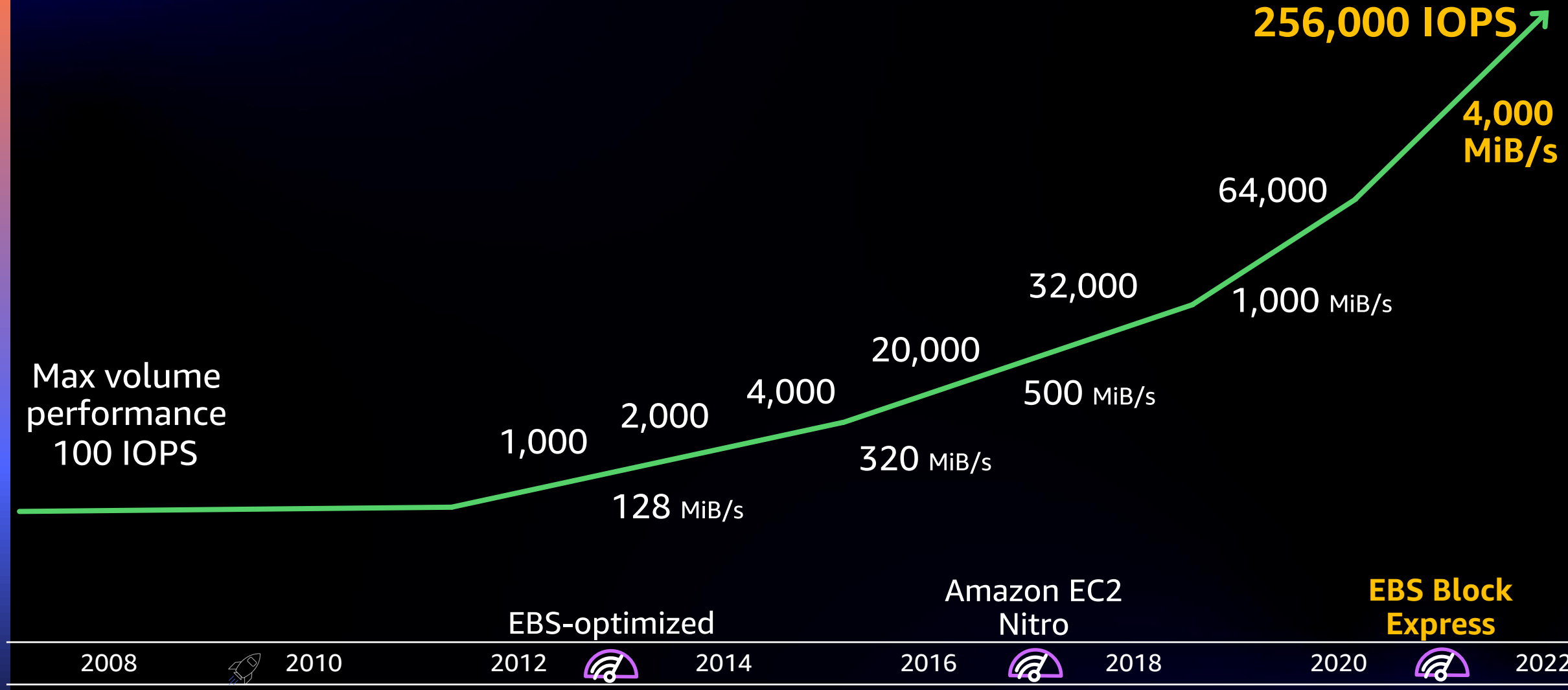
Multiple EBS volumes can be attached to an instance

Mix and match volume types for optimum performance

Use Elastic Volumes to change your volume characteristics

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Performance Innovations



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SAN in the Cloud

Overcome on-premises challenges



Scalable

Virtually unlimited scalability

Provision capacity that you need now with ability to grow on-demand



Simplify Management

Simplified provisioning of resources

Migrate data between storage tiers seamlessly to address changing workload requirements

Built-in security and recovery features that can be configured with a few clicks



Optimize Operations

Eliminate investments infrastructure management and complex datacenter operations

Eliminate the infrastructure lifecycle tax

No more disruptions and complex migration projects from EOL infrastructure

Seamless integration into AWS services

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SAN in the Cloud

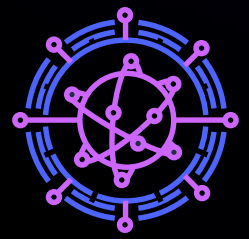
High Reliability

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SAN in the Cloud

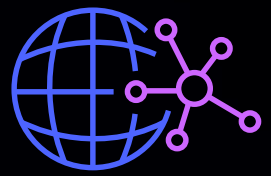
High reliability

Strong failure protection



Built-in redundancy within each volume's Availability Zone delivers superior protection against failure

High availability and durability



Keep your data safe with 99.999% availability and an annual failure rate (AFR) of 0.001%

Back up your data in Amazon S3



Store snapshots in Amazon S3, a cloud object storage service with 99.999999999% durability

Service-oriented Architecture for mission-critical workloads

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS is designed for...



99.999% service availability



Up to 99.999% durability

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

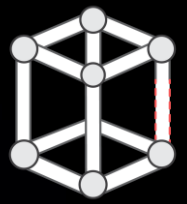
Blast radius reduction: Configuration manager

Region





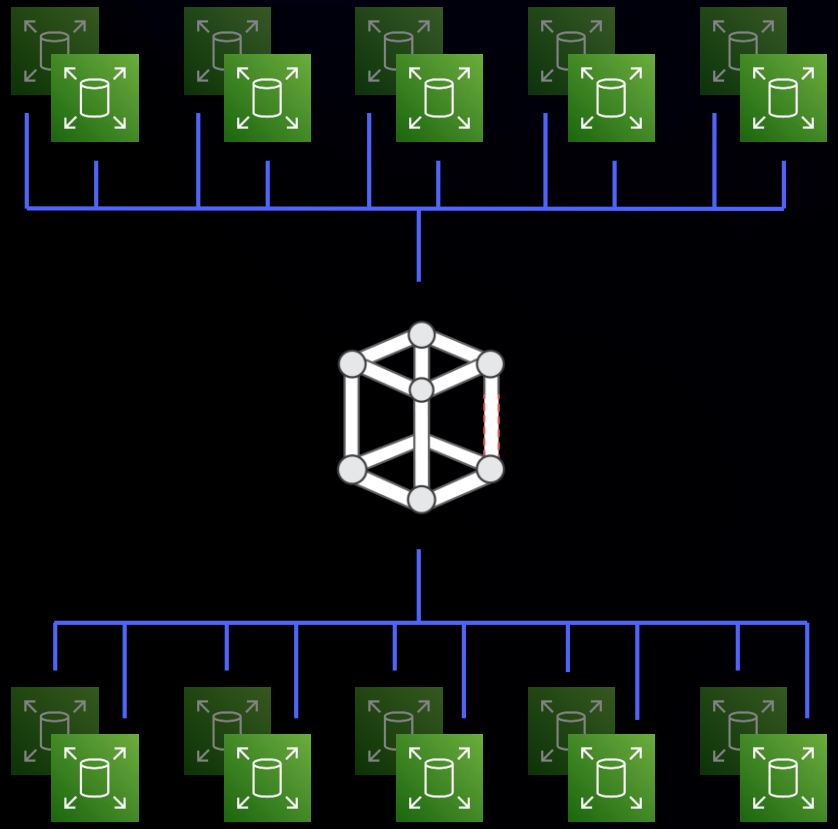
Availability Zone



aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

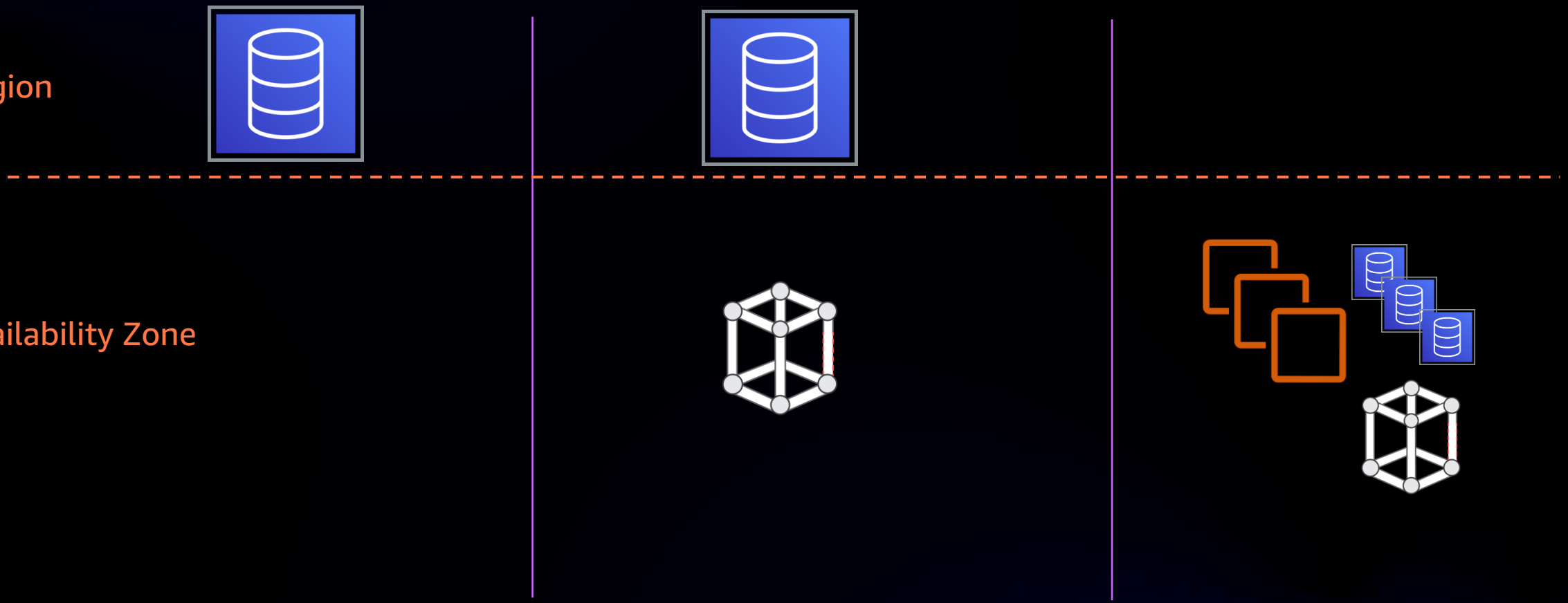
Configuration manager



- Authority map for host and volume
- Wholly contained in Availability Zone
- Standalone service
- Paxos-based replication

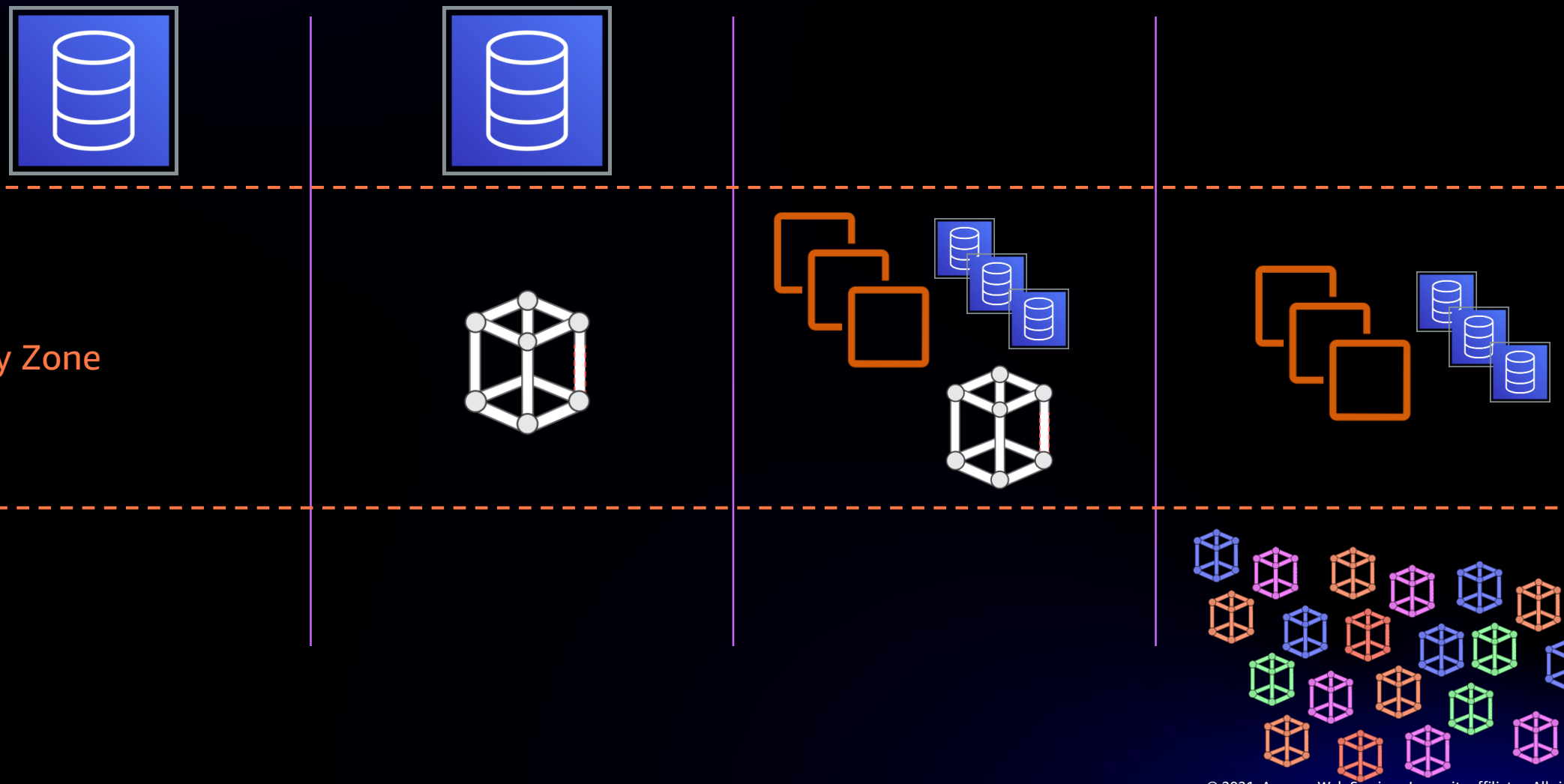
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blast radius reduction: AZ independence



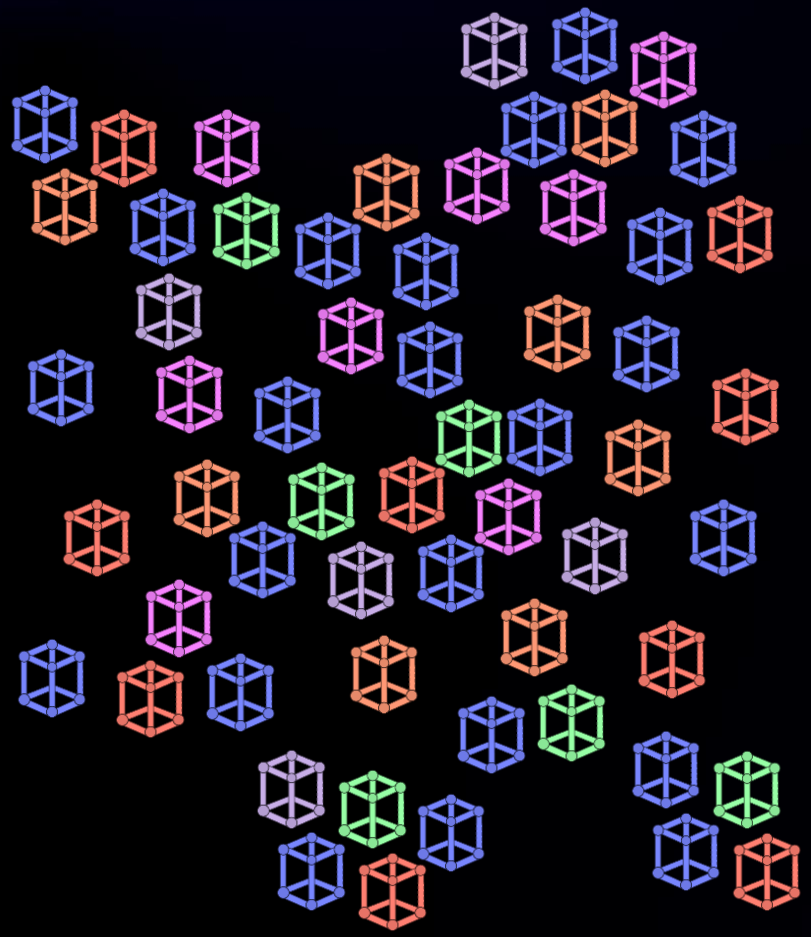
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blast radius reduction: Physalia



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Physalia: Millions of tiny databases



- Small number of clients per database
- Simple to manage
- Cell can reside near clients
- Reduced blast radius
- Each cell is a 7-node Paxos cluster

Read more: assets.amazon.science/c4/11/de2606884b63bf4d95190a3c2390/millions-of-tiny-databases.pdf

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Static stability is the ability for systems to return to steady state after small disturbances

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Protecting io2 Block Express Volumes

EBS Snapshots

Point-in-time copy of EBS volume

Incremental copies
Only changed blocks are copied

Stored in S3
11x9's durability

Crash consistent

Contains all data necessary to recover a volume

What are EBS snapshots



Snapshot use-cases

Backup data on EBS volumes

Meet RPO and RTO SLA objectives

Copy volumes within or across Availability Zones for redundancy

Copy volumes to another region for Disaster Recovery

Re-deploy production data for test/dev

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Protecting io2 Block Express Volumes

EBS Snapshots – features



Track

Tag snapshots at creation

Track and monitor snapshot resources

Enforce IAM rules based on tagging

“Cost allocation” tags provide visibility into snapshot storage costs



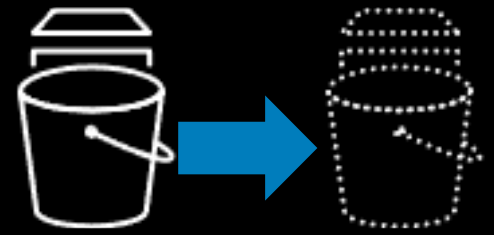
Secure

IAM policy driven privileges for snapshot management

Encrypted snapshots

Snapshots of encrypted volumes are encrypted at creation

Unencrypted snapshots can be encrypted during copy process



Share

Allow other accounts to copy a snapshot or create a volume from a snapshot

Enable test/dev activities

Enable re-deployment of custom AMIs

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Protecting io2 Block Express Volumes

EBS Snapshots – features



Track

Tag snapshots at creation

Track and monitor snapshot resources

Enforce IAM rules based on tagging

“Cost allocation” tags provide visibility into snapshot storage costs



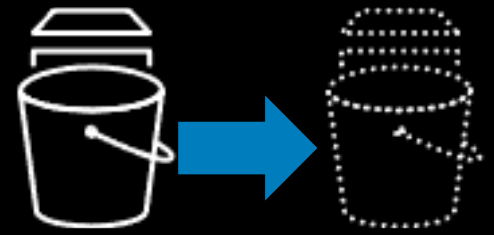
Secure

IAM policy driven privileges for snapshot management

Encrypted snapshots

Snapshots of encrypted volumes are encrypted at creation

Unencrypted snapshots can be encrypted during copy process



Share

Allow other accounts to copy a snapshot or create a volume from a snapshot

Enable test/dev activities

Enable re-deployment of custom AMIs

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Protecting io2 Block Express Volumes

EBS Snapshots – features



Track

Tag snapshots at creation

Track and monitor snapshot resources

Enforce IAM rules based on tagging

“Cost allocation” tags provide visibility into snapshot storage costs



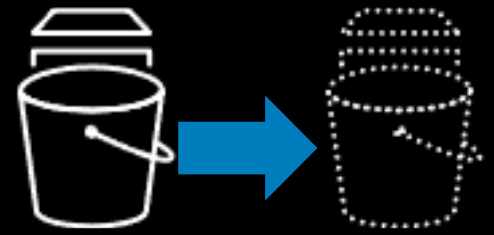
Secure

IAM policy driven privileges for snapshot management

Encrypted snapshots

Snapshots of encrypted volumes are encrypted at creation

Unencrypted snapshots can be encrypted during copy process



Share

Allow other accounts to copy a snapshot or create a volume from a snapshot

Enable test/dev activities

Enable re-deployment of custom AMIs

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SAN in the Cloud

Performance

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SAN in the Cloud

Target applications and workloads

Workloads that are sensitive to latency, IOPS, and throughput

IOPS

Hundred of thousands of IOPS

Latency

Sub-millisecond latency

Enterprise applications



SAP ERP, Oracle ERP
Microsoft SharePoint
Microsoft Exchange

Relational databases



MySQL, PostgreSQL
SQL Server, Oracle DB
SAP HANA, IBM DB2

NoSQL databases



Cassandra, MongoDB
CouchDB

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Understand your workload

Need more detail?

```
$ iostat -xdm /dev/nvme2n1 5
    :
Device:         rrqm/s   wrqm/s     r/s     w/s    rMB/s    wMB/s avgrq-sz avgqu-sz   ...
nvme2n1           0.00     0.00 24649.80 1564.80   995.53     6.11    78.25    40.31   ...
```

995.53 MB/s / 24650 reads/s ~= 40KiB / request

6.11 MB/s / 1565 writes/s ~= 4KiB / request

78 sectors / 512 B/sector ~= 39 KiB / request

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Understand your workload

Going deeper with blktrace

```
$ sudo blktrace -w 300 /dev/nvme2n1
  :
$ blkparse -d nvme2n1.bin -i nvme2n1 -o
$ btt -i nvme2n1.bin -B blk_offsets
$ ./bt_analyze.py blk_offsets_259,4_r.dat blk_offsets_259,4_w.dat
```

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Blktrace analysis

```
    with open(tracefile) as f:
        for line in f:
            start, end = line.split()[1:]

            start = int(start)
            end = int(end)

            num_ios += 1

            # keep track of sequential segments
            if start != prev_segment:
                num_segments += 1
            prev_segment = end

            # convert sectors to bytes
            sz = (end - start) / 2
            iosize[sz] += 1
            total_size += sz
```

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Understanding your mission

Going deeper with blktrace

Read

```
Sequential: 0.0006%
I/O by Size:
   0 - 4 KiB: 48
   4 - 8 KiB: 2
   8 - 16 KiB: 259776
   16 - 32 KiB: 65996
   32 - 64 KiB: 320934
   128 - 256 KiB: 2
```

Write

```
Sequential: 99.9879%
I/O by Size:
   0 - 4 KiB: 32973
```

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Building a better network protocol



- Throughput limited by incast
- Difficult to offload in hardware
- Slow to recover in failure

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Building a better network protocol



- Understand your environment
- Design for failure
- Recognize properties of your data stream

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SRD Storage Fabric



- Reliable high performance low-latency network transport
- Native multi-pathing enabled
- Congestion control for large-scale cloud deployments.
- High throughput; minimizes latency jitter
- Fast and transparent recovery from network failures

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Nitro card for EBS



NVMe controller

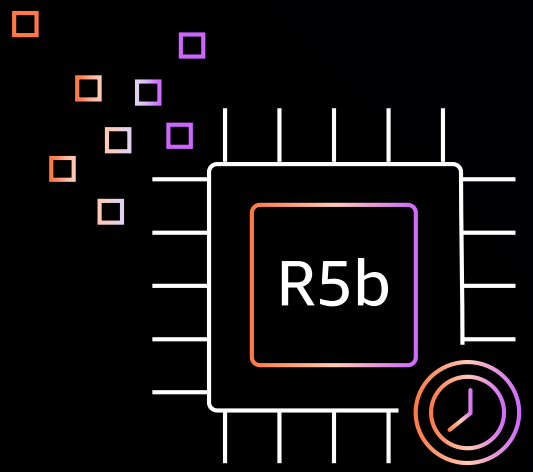
- Standard drivers broadly available

EBS data plane

- Encryption support
- NVM to remote storage protocol
- EBS optimized by default

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Instance family: R5b



- Same VCPU:memory ratio as R5
- Up to 260,000 IOPS
- Up to 60 Gbps EBS optimized BW (7,500 MiB/s)
- Built on Nitro + SRD

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Moving to services-based architecture improves control plane agility. Can we do the same in a high-performance data plane?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Disaggregated data plane

- Reducing code in the fast path improves performance

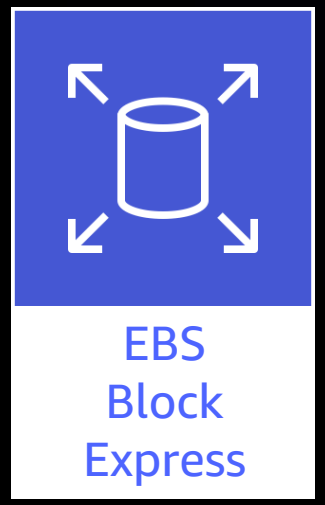
  Think critically about what absolutely needs to be there!

- Smaller code is easier to reason about security and correctness
- Can move faster with defined APIs

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS Block Express

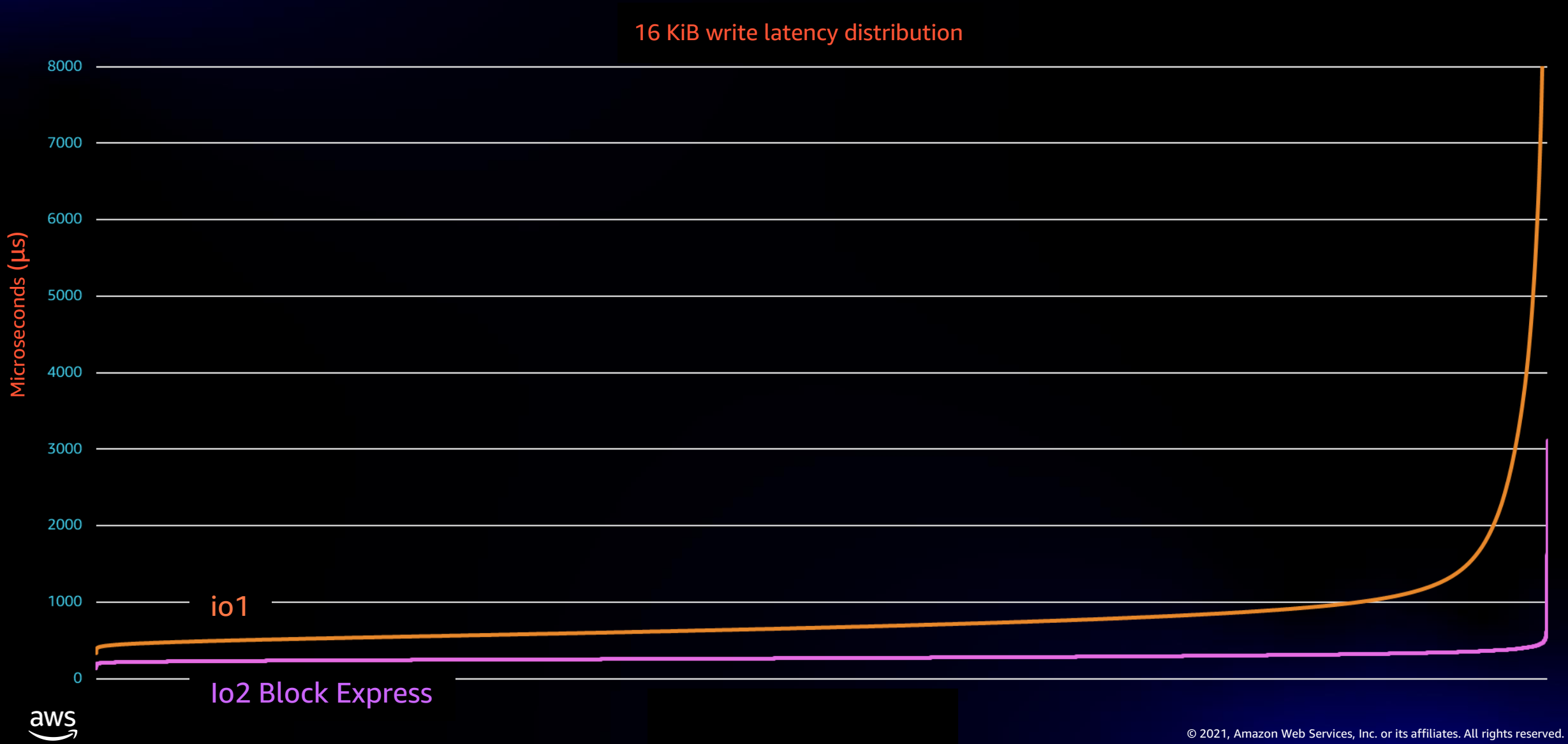


- Disaggregated infrastructure
- Disaggregated data plane and data services
- Nitro and SRD
- More consistent I/O latency

aws

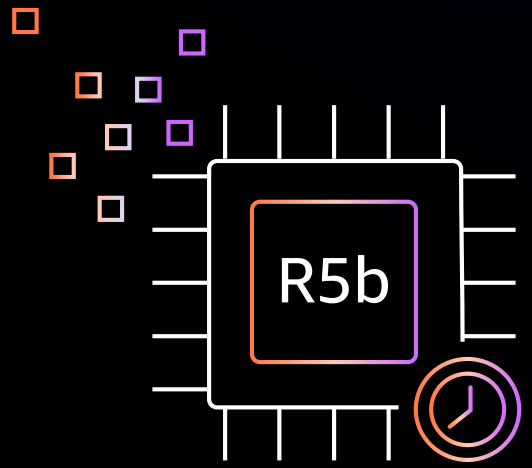
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Io2 Block Express: Performance

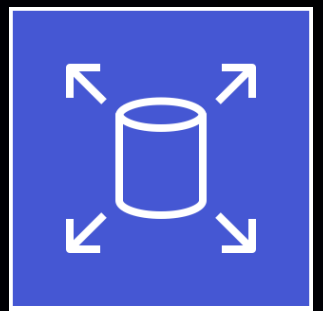


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

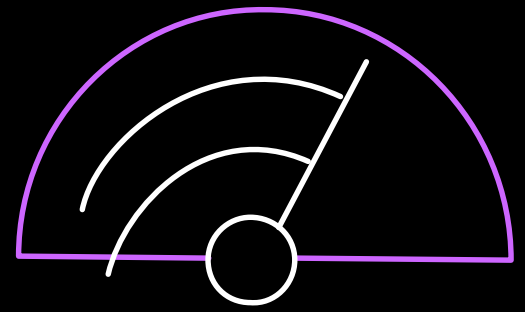
SAN in the cloud



+



=



3x–4x IOPS and Throughput

for high-performance applications

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SAN in the Cloud

Security

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SAN in the Cloud

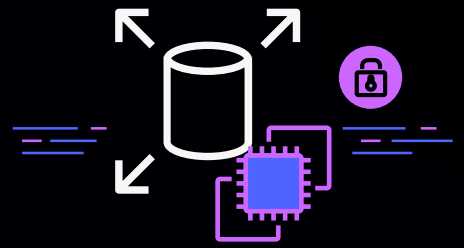
Built-in cloud security

Robust encryption



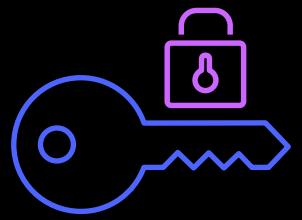
Protect data at-rest, data in-transit, and all volume backups with EBS encryption

Encryption by default



A single opt-in setting enables you to easily achieve compliance goals by encrypting all new volumes as they are created.

Simple and easy protection



Secure your data with ease, and with zero impact on performance. Encryption is supported by all volume types and it includes built-in key management.

SECURE FOR MISSION-CRITICAL WORKLOADS

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Encryption – Amazon EBS



EBS encryption: data volumes

- Integrates with Amazon Key Management Service (KMS)
- Uses Customer Master Keys (CMKs)
- AES-256 encryption for:
  - Data at rest inside the volume
  - Data moving between the volume and instance
  - Snapshots created from the volume
  - Volumes created from such snapshots

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS Encryption: Performance



EBS encryption: data volumes

Does choosing EBS Encryption reduce EBS Optimized Performance?

EBS Optimized Performance is the same with and without encryption for all Nitro instances

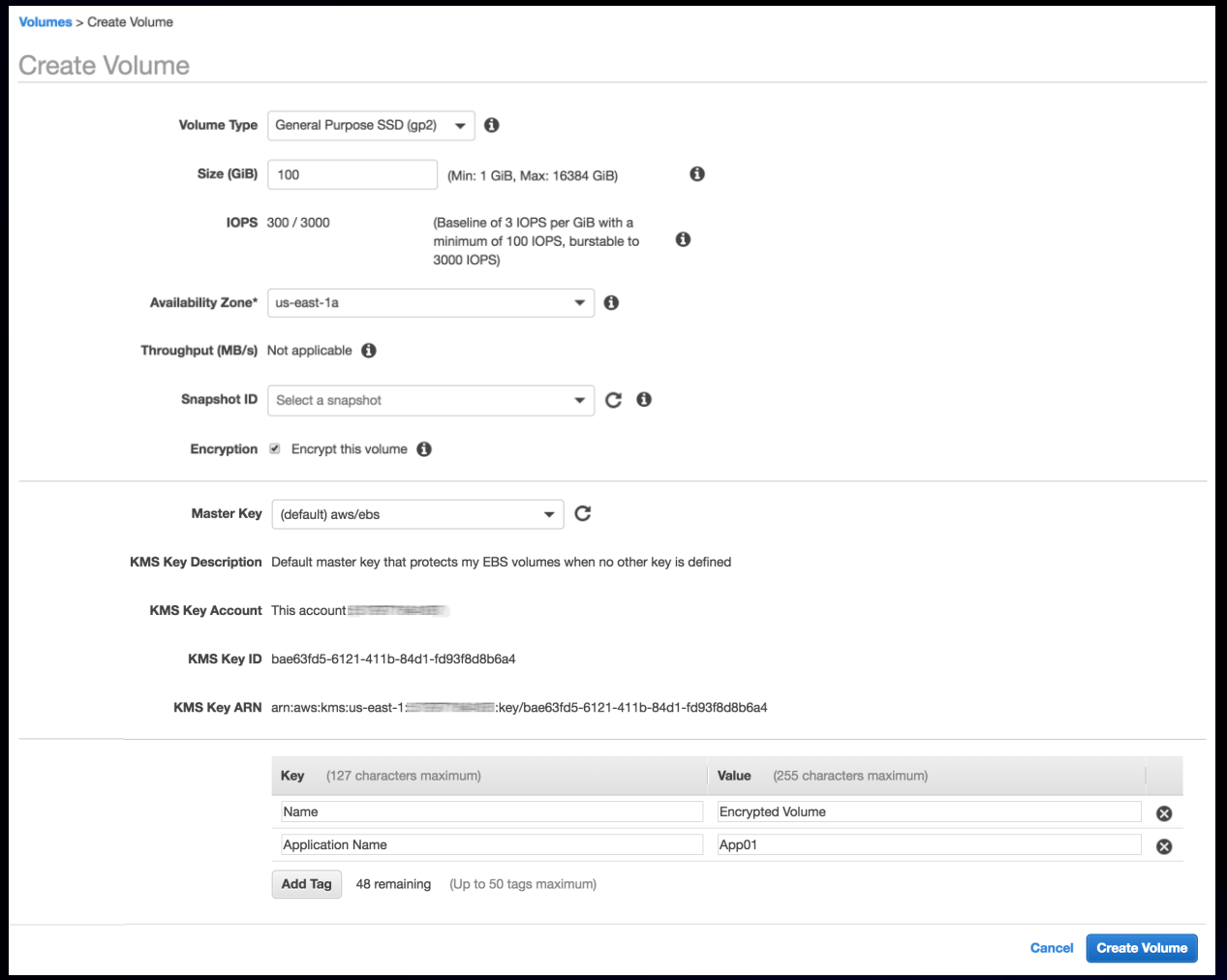
(and many previous generations)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS Volume encryption



EBS encryption: data volumes

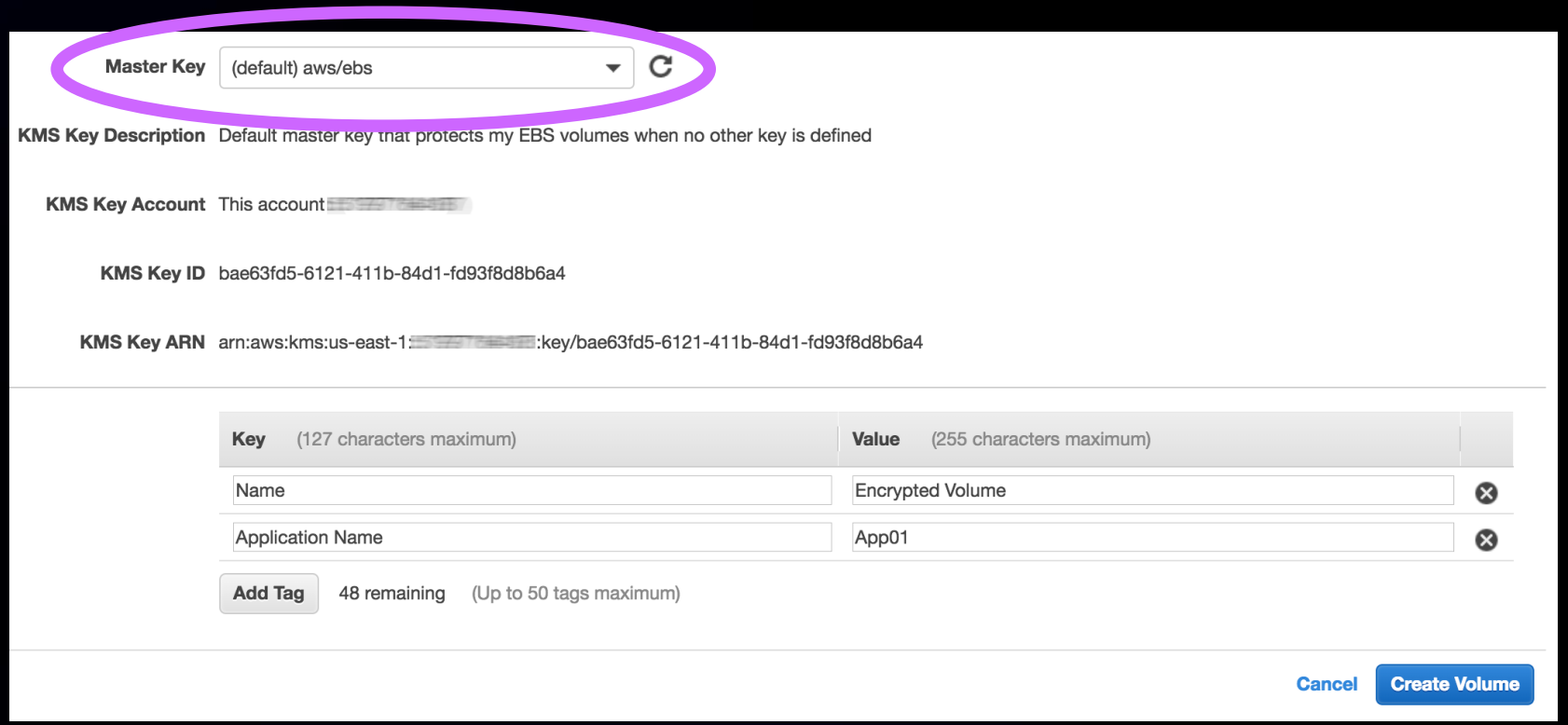


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS Volume encryption



EBS encryption: data volumes



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Previously ...



With IAM policies, you can prevent unencrypted volumes from launching

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

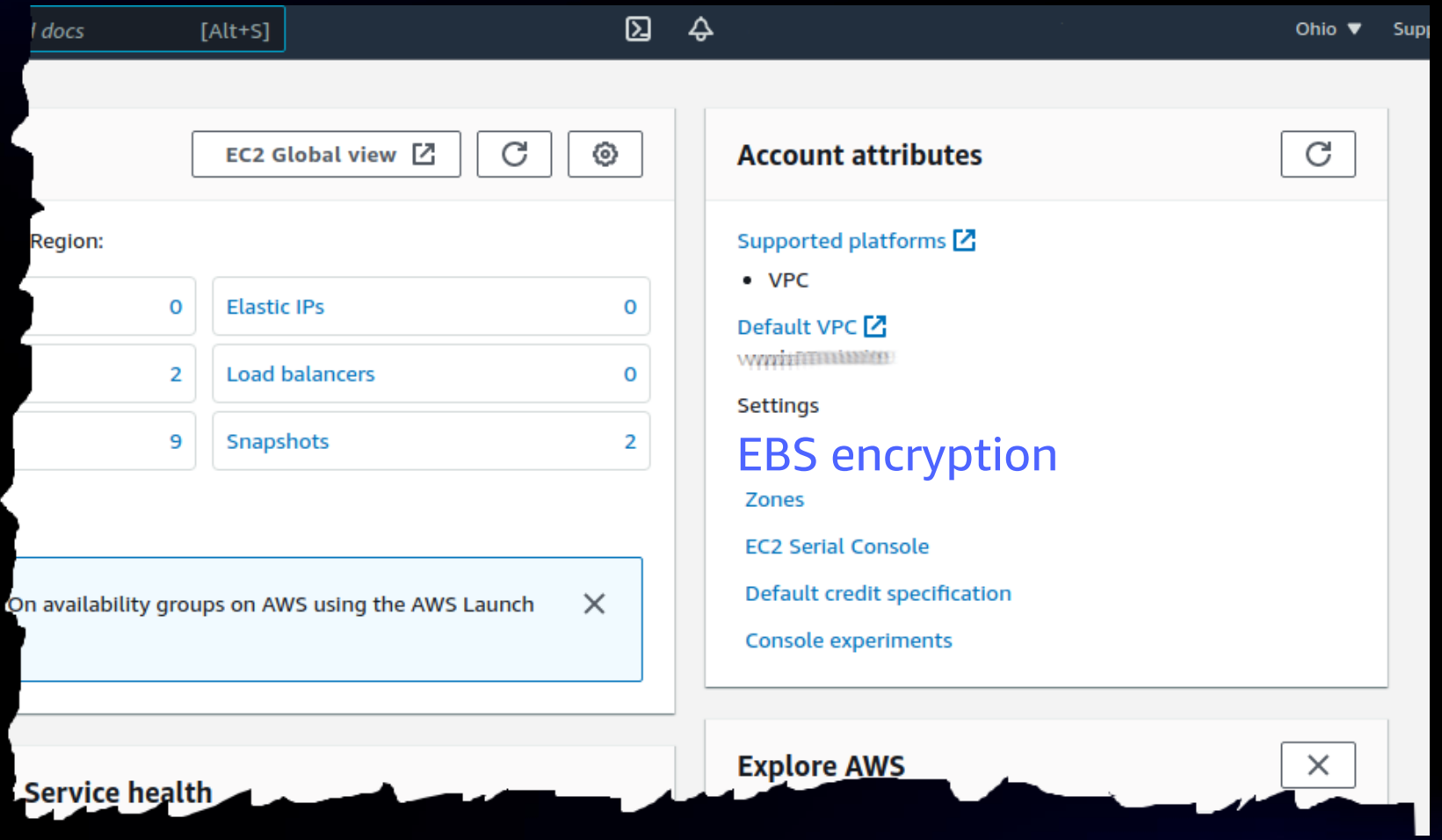
EBS encryption by default



Without change to workflows, newly launched volumes + snapshots are encrypted

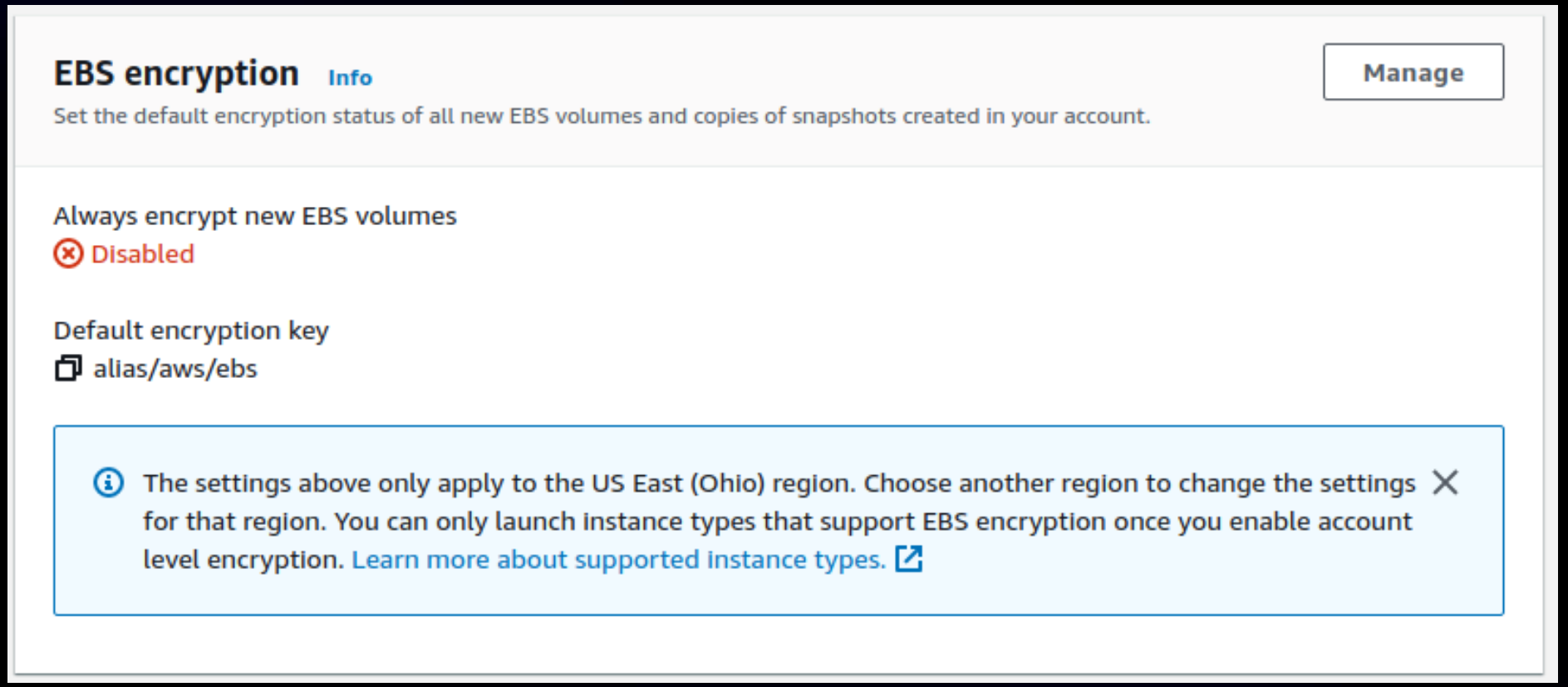
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS encryption by default



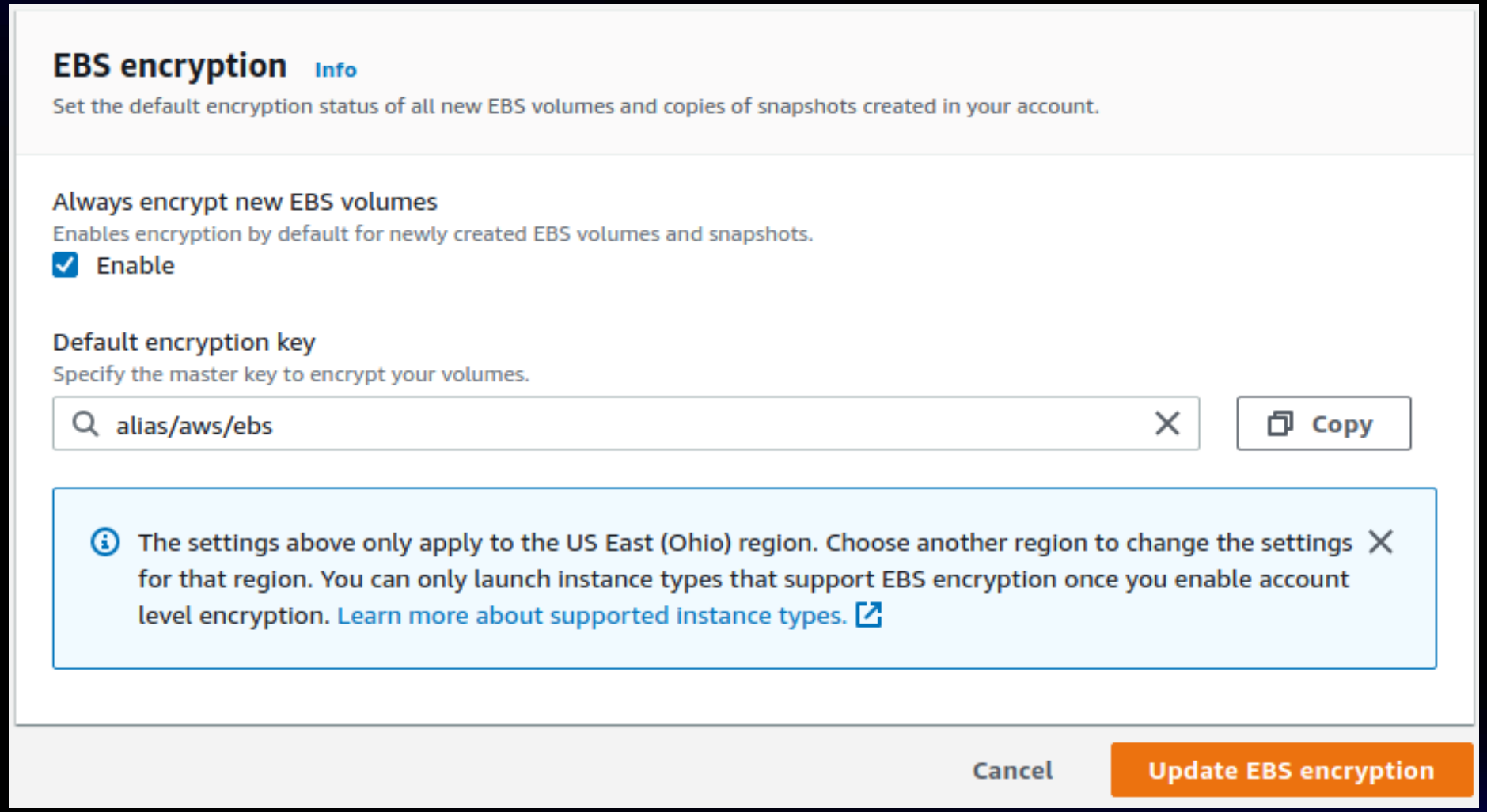
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS encryption by default



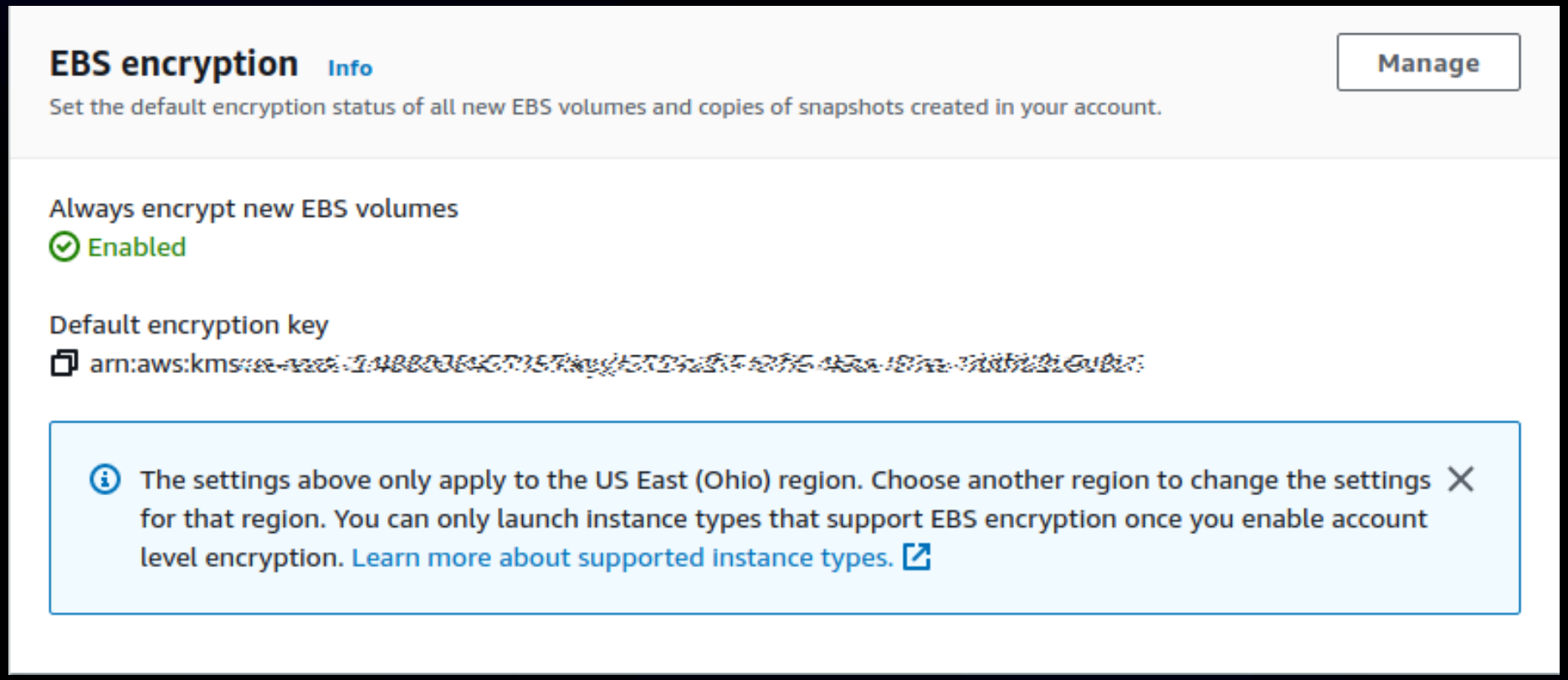
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS encryption by default



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS encryption by default



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

If you remember nothing else …



Monitor access

Keep tabs on which snapshots you are sharing and why



Account Level Encryption

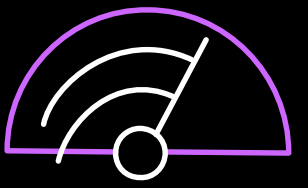
Encryption is a check-box – literally! Use it.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EBS Key takeaways



- Continue to **lower costs**
- Innovative **performance** and **reliability**





- **Flexible** tools to scale and manage
- **Simple** to use and manage **securely**



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Marc Olson

@molson

Prarthana Karmakar

@kprarthana

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.