# TR-4958: Best Practices for Electronic Design Automation Workloads with Amazon FSx for NetApp ONTAP

*Optimizing EDA and semiconductor workloads on Amazon Web Services (AWS)*

**First published Feb 10, 2023**

## Notices

# Contents

# Introduction

NetApp® ONTAP®-based storage has been a premier choice for customers running semiconductor workloads on-premises. With Amazon FSx for NetApp ONTAP (henceforth referred as FSx for ONTAP), AWS and NetApp have collaborated to bring feature-rich ONTAP software to customers with the operational ease of an AWS-managed service. FSx for ONTAP provides fully managed, shared storage in AWS Cloud with the popular data access and management capabilities of ONTAP. This document presents best practices for using FSx for ONTAP for Electronic Design Automation (EDA) and silicon design workloads. The document includes a short background on typical EDA-storage use cases and requirements.

EDA workloads are high performance computing (HPC) workloads that can see hundreds of thousands of compute cores running parallel jobs on a single project design tree. EDA workloads include front-end verification simulations running tools such as Synopsys VCS, Cadence Xcelium, or Siemens EDA ModelSim simulators to back-end chip-timing-analysis jobs like Synopsys Primetime or Fusion Compiler. There could be as many as twenty-five or more EDA design tools used for design specification; design capture; constrained random simulation; synthesis; Place and Route (P&R); power, performance, and area (PPA) optimization; floor planning; and chip finishing.

Faster time to market (TTM) at lower design cost is becoming critical for relevancy in the industry. Consumers expect smaller, faster, and cheaper devices. These factors are triggering quicker adoption of advanced silicon manufacturing technologies. Advanced chip designs are constantly demanding increased compute and storage for EDA workloads. Amazon Web Services (AWS), with its virtually infinite capacity and wide selection of compute and storage options, is a natural fit for EDA. EDA vendors like Synopsys, Cadence, and Siemens EDA are optimizing their EDA workflows to take advantage of cloud and AI techniques. Vendors are utilizing massively parallel computing resources to speed up job-completion time, reducing time to results, time to bug identification and fix, thereby realizing better quality tape-outs in shorter design cycles.

EDA storage solutions demand that you must deal with high metadata and highly parallelized access. This document provides configuration options for FSx for ONTAP to achieve the best performance at scale for EDA storage.

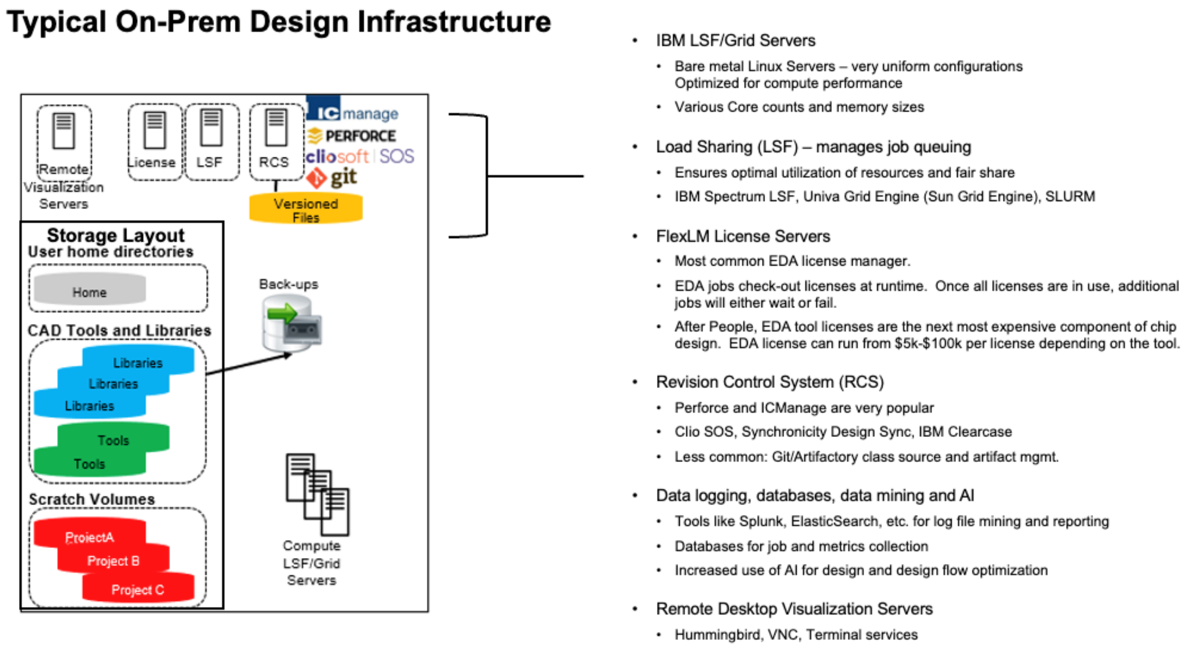
# EDA storage requirements

## Anatomy of an EDA cluster

Some of our largest customers use hundreds of thousands or more of parallel compute cores. High parallelism, high file count, and high metadata operations in EDA workloads make optimizing for scale-out performance critical. EDA is not a single workload. Sometimes more than fifty different tools can be used in a single semiconductor chip design project. These tools vary from providing functionality for design specification; verification to timing, power, and area optimization; and chip finishing.

NetApp has been a data management and storage leader in the semiconductor industry from its earliest days. On-premises, NetApp customers use NetApp A700s and A800 class all-flash storage appliances running the latest ONTAP storage operating system. Over the years, customer workloads have demonstrated that EDA workloads benefit from all-flash storage due to the ability drive highly concurrent operations at lower latencies. Customer workloads have also shown that the scale-out performance of NetApp FlexGroup volumes can dramatically improve the number of parallel jobs design teams can run, which in turn leads to a faster total turn-around time for EDA jobs. A typical on-premises EDA cluster is depicted in Figure 1.

**Figure 1: Typical on-premises EDA cluster.**




The cloud enables EDA design flows to scale compute with almost no limit using the latest compute cores available. Therefore, the underlying storage and data management system must be designed to support parallel cloud-scale workloads.

The recommendations in this document are focused on configuring FSx for ONTAP for maximum parallel compute performance while maintaining latency suitable for EDA. Customers can cost-optimize their deployments by right-sizing the IOPS and throughput and enabling intelligent data tiering.

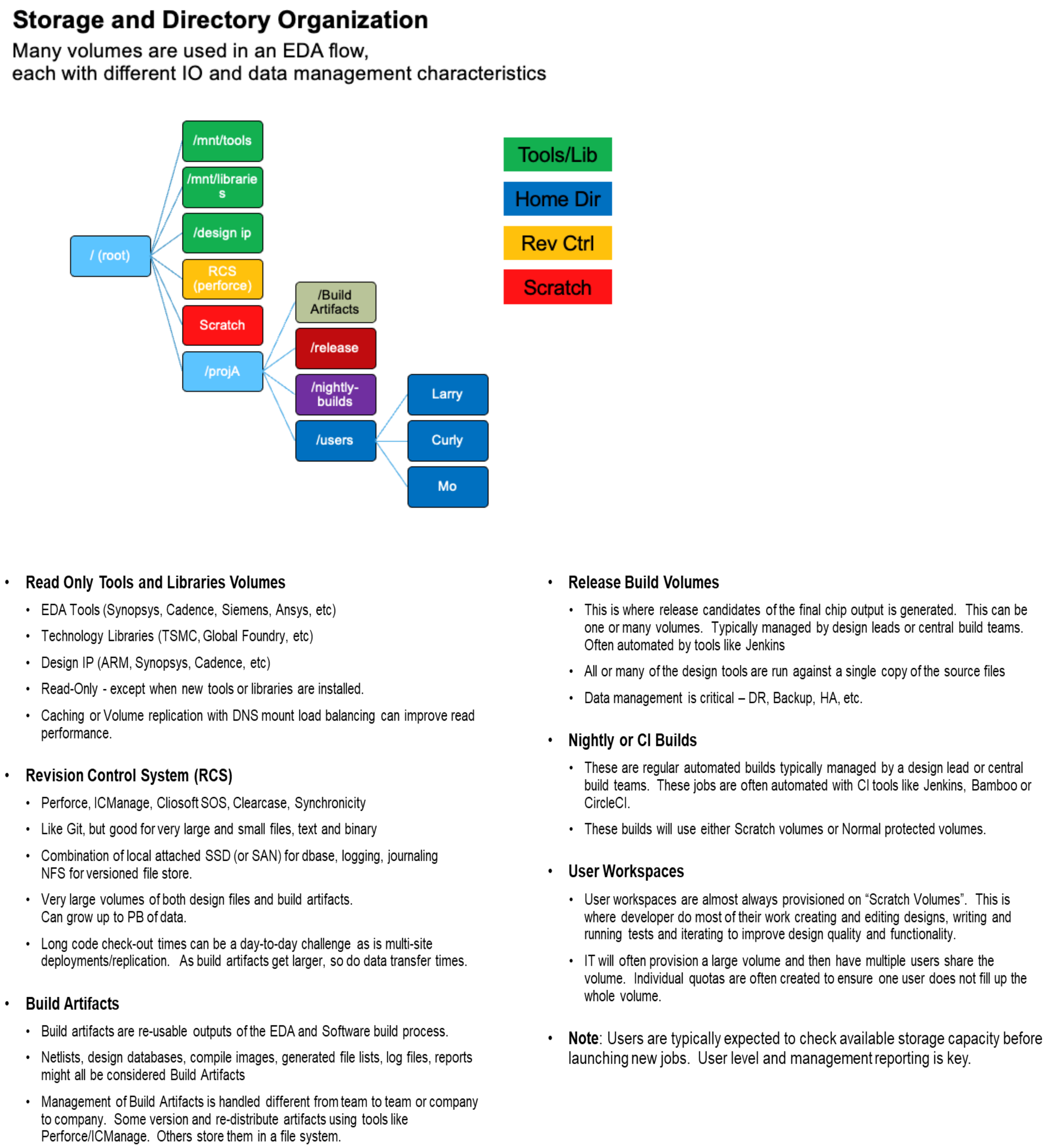
## EDA volumes and directories

EDA workflows mount and use data from many different volumes in a single workflow. Not all data volumes have the same read-write characteristics. Knowing how each volume is used in an EDA flow can provide opportunities for runtime performance and utilization optimization.

For example, `/mnt/tools`, `/mnt/libs`, or `/mnt/libraries` are common storage volume names representing the location of the EDA tools and design libraries used in a design flow. Within these volume(s), there are often tens to hundreds of tools and tool versions installed. Tools and libraries are often installed, centrally managed by CAD teams, and then shared with multiple teams and projects.

The typical directories in an EDA environment, along with description of contents and function of each, is as show in Figure 2.

**Figure 2: Directory organization in EDA.**




IT storage teams typically provision storage based on the needs of business units, teams, projects and/or by specific design flows. Each company’s design flows are unique and yet similar in various ways. With FSx for ONTAP, customers get the agility and scalability of the cloud and only pay for provisioned storage. Customers do not need to make upfront investments as is the case for on-premises, which can lead to stranded capacity if business objectives change.

## EDA data types

### Tool and library volumes - Optimized for read performance

EDA tools like those from Synopsys, Cadence, and Siemens EDA are typically installed in tool directories mounted like `/mnt/tools/synopsys`, `/mnt/tools/cadence`, and so on. Each of these installation directories can contain many different tools and tool versions.

Similarly, third-party design technology libraries are typically installed in directories mounted like `/mnt/libs/tsmc`, `/mnt/libs/arm`, `/mnt/libs/dolphin`, and so on. These libraries contain either design IP (like a USB or CPU core) or EDA vendor libraries from device manufactures like TSMC, Intel, Global Foundries, UMC, or Samsung, which describe the physical power, area, and performance data associated with the transistor or cell representation of the process node. Tech libraries typically contain thousands of files containing an ever-increasing amount of data.

Internal libraries or design IP are another class of data used in designs. These items might be subsystems or blocks developed by one team within the company and delivered to another team for inclusion in a full chip design. For example, an ARM core designed by the processor team might be delivered and used by a team to integrate the core into a mobile device.

The key characteristic of tools and libraries is that the data is typically read-only or read-mostly. Writes occur when new versions of tools or libraries are installed, but seldom are tools and libraries over-written. New tool or library versions are typically installed next to the prior versions.

As such, we can take advantage of this by optimizing the infrastructure for read performance. This might include NFS mount options to maximize compute server data caching to improve I/O latency.

Tools and libraries are accessed by the EDA design flows and might see very high concurrent accesses from the 10k-100k parallel jobs simultaneously running in the compute farm. As such, I/O performance and scale is critical. Tools like Synopsys PrimeTime, a timing analysis tool, loads thousands of library files simultaneously from thousands of servers resulting in high read loads, which might cause performance bottlenecks. The reference architecture proposed in this document addresses how to use FlexCache to improve read performance by load-balancing NFS mounts across multiple FlexCache tool and library volumes. Figure 3 depicts the these directories.

**Figure 3: Tool and library directory structure in EDA.**

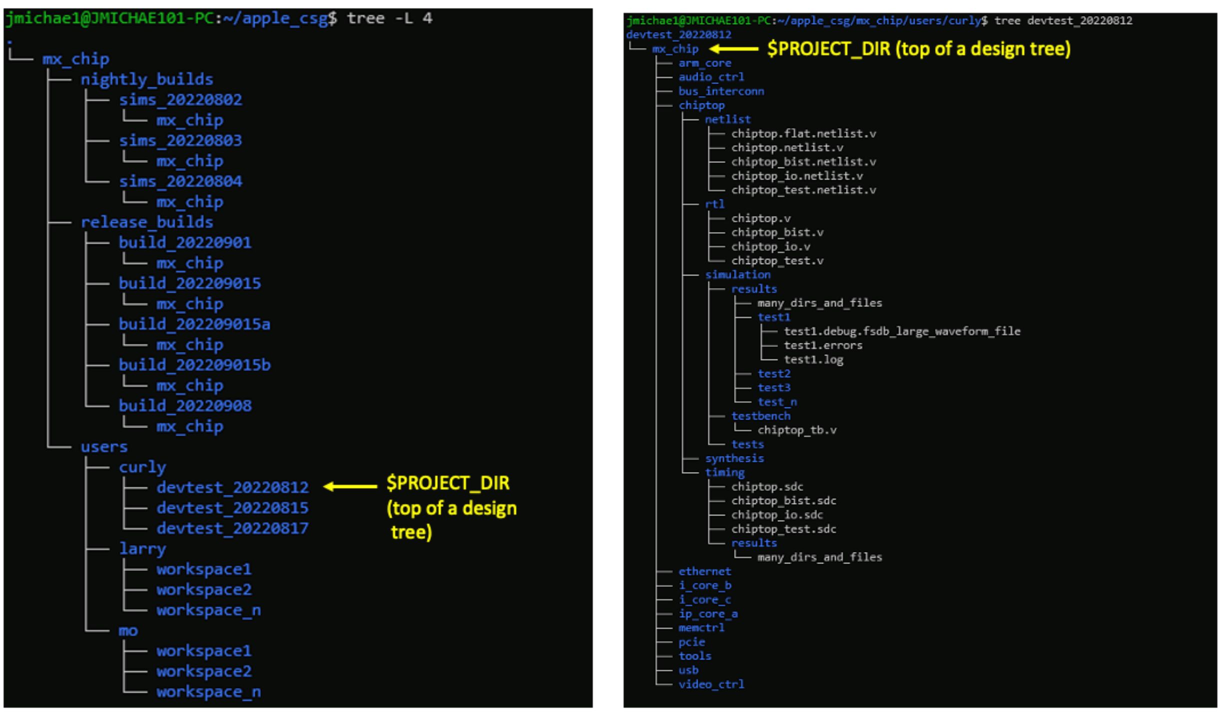
## Scratch volumes – Optimized for maximum read-write performance

Scratch volumes are used for user workspaces, nightly regressions, CI/CD build flows, and many other development purposes. The term scratch comes from the idea that the data is temporary or has such a high change rate that traditional back-up and archiving is unnecessary.

Scratch is where active development is happening. Engineers launch EDA jobs and wait for them to complete so that they can check to see if a test passed or failed after a design change or a new functional test. For example, designers might be waiting for the results of thousands of jobs to check the power performance and area (PPA) characteristics of the chip after a functional change or optimization. Performance is critical, and both storage capacity and compute scale are critical attributes.

Scratch volumes should be optimized for maximum read-write performance. Modern chip design flows can use as many as one hundred different tools with a wide range of I/O profiles (read-to-write ratios, big versus small files, sequential versus random access) and can vary dramatically. These environments are often shared by multiple flows and multiple teams. EDA scratch area snapshots are shown Figure 4.

**Figure 4: Typical EDA scratch space directory hierarchy.**




## Build artifacts

Build artifacts can be many types of data generated from the EDA or software development workflow. A chip or IP netlist generated from Register Transfer Level (RTL) code using Synthesis tools is a build artifact. A generated gate level netlist might be reused by the verification team for gate-level simulations, reused by the emulation team, and reused by the timing analysis and floor planning team to name a few. Modern semiconductors are made up of both soft, firm, and hard IP that is developed by one team and delivered to another team for integration into a design. The artifacts of an IP or subsystem-development team might become reused IP deliverables provided to another team for integration.

Companies manage and share build artifacts differently. Some teams use design data management tools for revision control and intellectual property (IP) management tools to maintain build artifacts versions and distribute IP to teams and development sites. Other companies distribute build artifacts by replicating data volumes or directories. Others use symbolic links to shared files and directories into development areas. Build artifact volumes are typically designed for balanced read-write. Many NetApp customers use data replication technologies like NetApp SnapMirror™ or FlexCache® technology to distribute artifacts across multiple sites.

## EDA performance

EDA workloads demand high IOPS and throughput from shared network storage. The more performant a file system is, the more EDA jobs can be run in parallel and the faster design teams can find bugs, iterate on design optimization, and so on. The sizing goal is to optimize the configuration for the maximum possible IOPS and throughput.

EDA workloads are HPC workloads the require maximum single-job runtime performance (the shortest possible wall-clock time to job completion). Modern EDA workflows have 20k to 100k compute cores running in parallel on a single design (composed of a set of volumes), therefore optimizing FSx for ONTAP to maximize the number parallel cores is critical. Storage I/O performance should be carefully monitored as the number of parallel jobs is increased, which can be observed by the latency rapidly increasing above 3ms; this is observed as the knee of the performance curve.

### Single job performance measured by wall-clock time

EDA job performance is typically measured by the wall-clock time (the duration from the start and end of a job as measured by an end-user). How long did it take to run a single chip-design simulation? How long did it take to complete a timing analysis or floor plan?

An improvement in job runtime is critical because it improves the productivity and utilization of design engineers and EDA licenses. Engineers and EDA licenses make up over 90% of chip development costs. Wall-clock improvements of 5 to 10% are considered highly significant and are worth spending money on. EDA infrastructure costs are approximately 10% of chip development costs and storage is roughly 4% of the IT spend, so spending more on storage performance pays for itself if it improves the productivity and utilization of engineers and licenses.

### Performance measured by total turn-around time of all jobs

EDA performance is also measured in total turn-around time (TTT). This is the time it takes for a collection of jobs to complete. For example, how long does it take for one thousand simulations to run start to finish? The answer to that question is highly dependent on whether the job tests were run serially, in parallel, or in some combination of the two. Load-sharing tools like LSF, Grid Engine, or Slurm allow engineers to submit large sets of jobs to run in parallel. The more jobs that can run in parallel, typically the faster all jobs complete.

**Table 1: Running 20k jobs.**

| | Total jobs | Maximum parallel jobs | Average job runtime (min) | TTT (min) | TTT (hrs) |
|---|---|---|---|---|---|
| Example 1 | 20,000 | 1,000 | 10 | 200 | 3.33 |
| Example 2 | 20,000 | 5,000 | 10 | 40 | 0.67 (5x faster) |

As you can see from Table 1, running 5,000 jobs in parallel results in a 5x improvement in job turn-around time, which in turns means 5x faster for bug discovery, 5x faster for design closure iterations, and 5x more productive engineers.

Optimizing for maximum parallel performance and scale is key when considering EDA design architectures. Cloud computing enables unlimited parallel compute scale because the cloud is vast and elastic relative to the fixed resources in traditional datacenters. And in turn, the cloud provides significant improvement in TTT and productivity.

The ability to run more jobs in parallel is as important a consideration as optimizing single-job wall clock time performance.

# FSx for ONTAP

## FSx for ONTAP introduction

Amazon FSx for NetApp ONTAP is a fully managed service that provides highly reliable, scalable, high-performing, and feature-rich file storage built on the popular NetApp ONTAP file system. FSx for ONTAP combines the familiar features, performance, capabilities, and API operations of NetApp file systems with the agility, scalability, and simplicity of a fully managed AWS service.

FSx for ONTAP provides feature-rich, fast, and flexible shared-file storage that is broadly accessible from Linux, Windows, and macOS compute instances running in AWS or on premises. FSx for ONTAP offers high-performance solid-state drive (SSD) storage with sub-millisecond latencies. A FSx for NetApp ONTAP file system has two storage tiers: primary storage and capacity-pool storage. Primary storage is provisioned, scalable, high-performance SSD storage that's purpose-built for the active portion of your data set. Capacity pool storage is a fully elastic storage tier that can scale to petabytes in size and is cost-optimized for infrequently accessed data. On each volume, data is automatically tiered to the capacity pool storage when automatic data tiering is enabled on a volume. The volume's tiering policy determines when data is tiered off to capacity pool storage. With FSx for ONTAP, you can achieve SSD levels of performance for your workload while paying for SSD storage for only a small fraction of your data.

Managing your data with FSx for ONTAP is easier because you can snapshot, clone, and replicate your files with the click of a button. In addition, FSx for ONTAP automatically tiers your data to lower-cost, elastic storage, reducing the need to provision or manage capacity.

FSx for ONTAP also provides highly available and durable storage with fully managed backups and support for cross-region disaster recovery. To make it easier to protect and secure your data, FSx for ONTAP supports popular data security and antivirus applications.

For customers who use NetApp ONTAP on-premises, FSx for ONTAP is an ideal solution to migrate, back up, or burst your file-based applications from on-premises to AWS without the need to change your application code or how you manage your data.

As a fully managed service, FSx for ONTAP makes it easier to launch and scale reliable, high-performing, and secure shared file storage in the cloud. With FSx for ONTAP, you no longer have to worry about the following issues:

- Setting up and provisioning file servers and storage volumes
- Replicating data
- Installing and patching file server software
- Detecting and addressing hardware failures
- Managing failover and failback
- Manually performing backup

FSx for ONTAP also provides rich integration with other AWS services, such as AWS Identity and Access Management (IAM), Amazon WorkSpaces, AWS Key Management Service (AWS KMS), and AWS CloudTrail.

## FSx for ONTAP – Deployment types

FSx for ONTAP can be deployed in single-availability zone (AZ) or in multi-AZ configurations.

The traditional EDA/semiconductor datacenter contains compute and storage in the same building. If there is a failure in the datacenter, compute and storage can be affected. This might be a result of a network outage, power outage, or any other issue. The EDA data stored on the NFS filers are typically not replicated in real time to secondary datacenters. Some EDA volumes might be replicated or backed up to another datacenter for disaster recovery several times per day, but not continuously.

The FSx for ONTAP single-AZ deployment type is analogous to a highly available pair deployed in a traditional datacenter as was described above. Single-AZ FSx for ONTAP file systems replicate your data and offer automatic failover within a single AZ.

The FSx for ONTAP multi-AZ deployment type on the other hand is more like a NetApp MetroCluster™ instance. MetroCluster instances are two NetApp filers in two different datacenters separated by a distance (in the same city or in an adjacent city). Every write that comes in on one filer is immediately sent to the second filer. Only after both filers acknowledge the write is the acknowledgement sent back to the computer that initiated the write. MetroCluster is typically used when there is no tolerance for outages; failover from one datacenter to another must be seamless, immediate, and completely fool proof.

Multi-AZ file systems provide added resiliency by replicating your data and supporting failover across multiple AZs within the same AWS region. Because writes are committed across AZs, multi-AZ file systems have slightly higher latencies when compared to single AZ file systems.

## FSx for ONTAP sizing considerations for EDA workloads

Modern EDA workloads utilize 20,000 to 100,000 parallel compute cores on a single design. High parallelism, high file count, high metadata workloads make optimizing for scale-out performance critical.

There is no single EDA workload. As many as 50+ different tools might be used to design chips, from the early design specification, to design, timing, power, and area optimization and chip finishing. These recommendations are focused on configuring FSx for ONTAP for maximum IOPS, minimum latency, and maximum parallel compute performance.

Infrastructure cost is always a consideration, but the productivity of engineers and the use of expensive EDA design tool licenses are a much larger expense than IT and cloud infrastructure costs. That said, after performance requirements are understood, then thr performance to cost trade-offs can be assessed.

## FSx for ONTAP disk selection and throughput requirements

All-flash (SSD-based) disk performance has demonstrated its value for demanding EDA and HPC workloads. Sub-millisecond latencies have improved performance of EDA workloads and have almost completely removed the need for performance-tuning storage as was common practice for 10k SAS and SATA drives.

SATA drives are still common in EDA, but only for backup and cold data storage. Active workloads are almost exclusively on high performance all-flash storage platforms on-premises. The same is true for cloud-based NFS storage. In a do-it-yourself NFS storage solution on AWS, pairing disk storage throughput performance to match the EC2 instance throughput is critical for maximizing workload performance.

FSx for ONTAP eliminates the operational undifferentiated heavy lifting by providing NetApp ONTAP as a fully managed AWS managed service. It provides feature-rich, fast, and flexible shared-file storage in two tiers: a primary pool for performance and a capacity pool that is cost optimized for infrequently accessed data.

Amazon FSx for NetApp ONTAP provides sub-millisecond file-operation latencies with SSD storage, and tens of milliseconds of latency for capacity pool storage. In addition, Amazon FSx has two layers of read caching on each file server—NVMe drives and in-memory—to provide even lower latencies when you access your most frequently read data. Each Amazon FSx file system provides up to multiple GBps of throughput and hundreds of thousands of IOPS. These features make FSx for ONTAP a great solution for EDA on AWS.

# EDA storage architecture patterns on AWS

While not a comprehensive list, EDA architectural patterns broadly fit into the following three categories:

- Full chip design on AWS
- Burst capacity onto AWS
- Backup and disaster recovery (DR) on AWS

## ONTAP data management - Full chip on AWS

To design a full chip on AWS, primary volumes, snapshots, and backups all reside on AWS. In addition to provisioning primary volumes on FSx for ONTAP, native snapshot and backup capabilities can be used to architect for disaster recovery (DR) and business continuity.

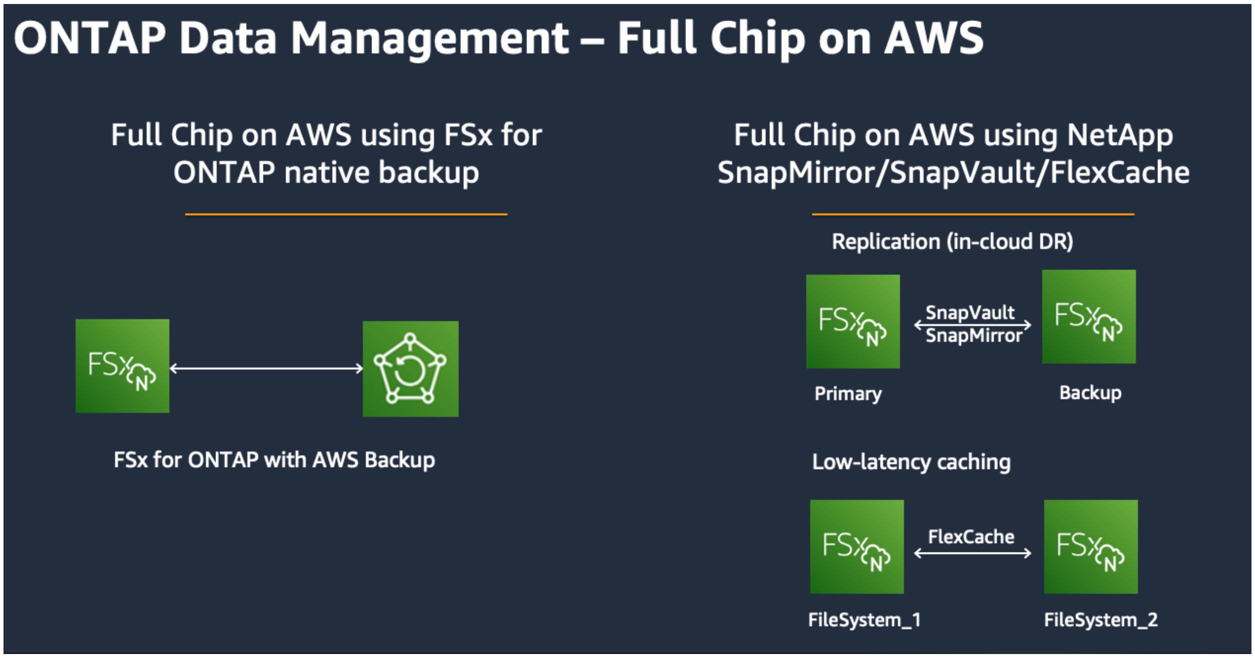
Furthermore, the NetApp SnapMirror and SnapVault data replication tools are also available out of the box for FSx for ONTAP. This feature, already familiar to many customers, allows replication relationships to be established across different AWS regions if needed.

Business continuity requirements, the recovery time objective (RTO) and the recovery point objective (RPO) determine the best approach to architecting DR for full chip projects on AWS.

NetApp FlexCache technology can be used to lazy load active data on demand onto a FSx for ONTAP file system from another file system. The feature enables performant cross AWS-region EDA. The methodology can also be used across an AZ or within the same AZ to horizontally scale file systems. A FlexCache instance can be created from primary or DR volumes.

The architectural patterns for full chip on AWS are depicted in Figure 5.

**Figure 5: ONTAP data management for full chip on AWS.**




## ONTAP data management - Burst capacity onto AWS

### Hybrid EDA – Burst compute to AWS

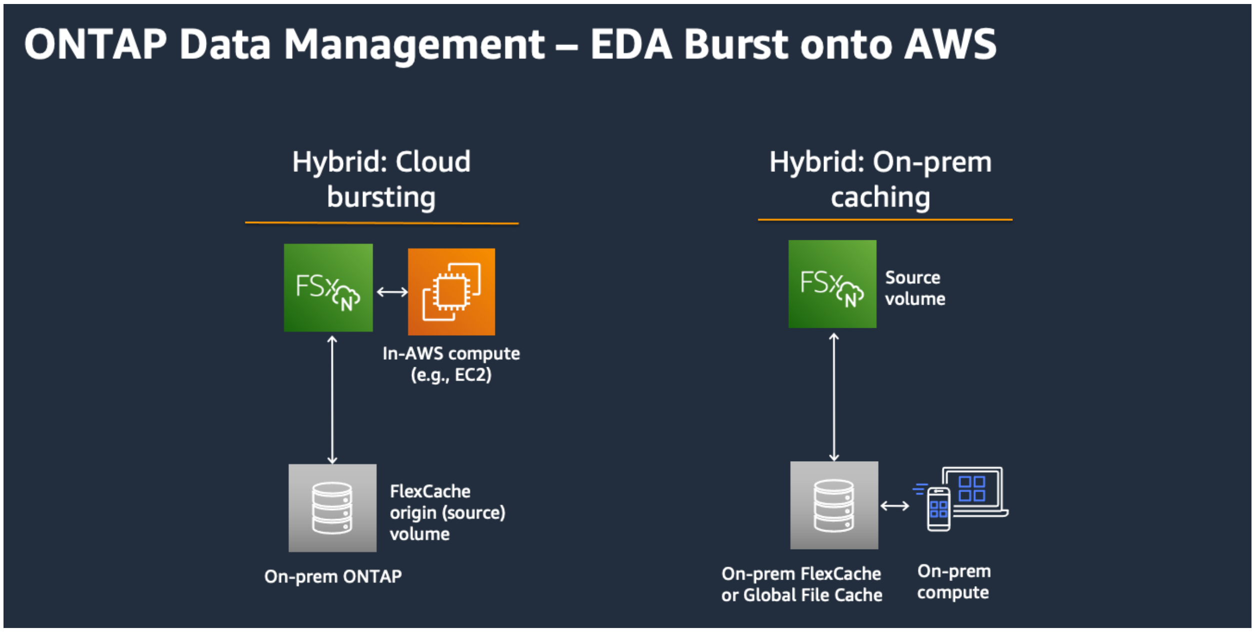
Several customers have used AWS to serve EDA burst capacity. As a complement to on-premises compute, essentially infinite compute capacity on AWS has made it possible for EDA customers to enable faster chip tape-outs. In this architecture data located on premises can be cached remotely by an FSx for an ONTAP file system using FlexCache. An example of this would be an EDA tools library stored on premises in Austin, TX, that remote users in Singapore could access with low latency. By using a FlexCache volume on an FSx for ONTAP file system in Singapore linked to the origin volume in on-premises, customers can burst their workflow to use compute in Singapore to provide low-latency access to cached data for end users. FlexCache supports both NFS and SMB file caching. Files can be cached between ONTAP file systems if desired.

### Hybrid EDA - Storage on AWS with on-premises caching

Although most customers prefer to migrate compute and storage to AWS, the inverse of the cloud bursting use case we just reviewed is also an architecture made possible with FSx for ONTAP. In this model, the tools library resides on an FSx for ONTAP file system. Using a FlexCache volume on premises linked to the source volume on an FSx for ONTAP file system, customers can cache data on premises for low-latency access. This architecture could be useful for sharing EDA directories (for example, tools and library releases) across geos with AWS when connectivity between on-premises sites does not exist or does not enable easy sharing.

Architectural patterns discussed above for modeling EDA burst on AWS are depicted in Figure 6.

**Figure 6: ONTAP data management for EDA burst onto AWS.**




## ONTAP data management – DR and archive on AWS

NetApp ONTAP is widely deployed as an on-premises EDA storage solution. With FSx for ONTAP, customers can use already familiar ONTAP features for data replication. Therefore, storing EDA backup, DR, and archive data on cost-efficient FSx for ONTAP is a popular starting point for EDA on AWS. ONTAP SnapVault software replicates NetApp Snapshot™ copies from the on-premises NetApp storage array to the FSx for ONTAP file system for longer term retention on the destination. SnapMirror replicates the active file system and Snapshot copies for DR purposes. Data can also be replicated across AWS regions or AZs with SnapVault and SnapMirror. DR architecture on AWS is depicted in Figure 7.

**Figure 7: ONTAP data management for EDA burst onto AWS.**

## Multi-site EDA architectural patterns

FSx for ONTAP helps to realize complex EDA architectures with relative ease. For the purposes of this document, a site can be a single on-premises data center facility or a single AWS region. Therefore, a multi-site architecture could contain a combination of one or more on-premises data centers and one or more AWS regions. Figures 8 and 9 depict multi-site EDA architectures using AWS for multi-project and multi-team organizations.

**Figure 8: Multi-site AWS FSx for ONTAP with on-premises architecture.**

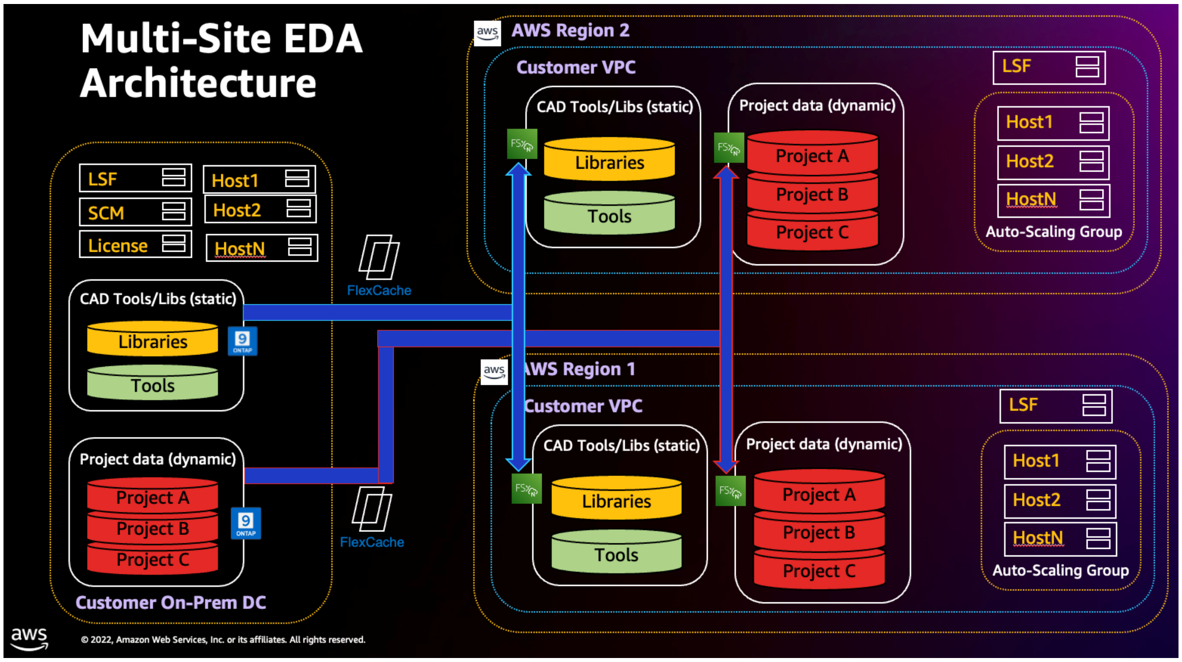
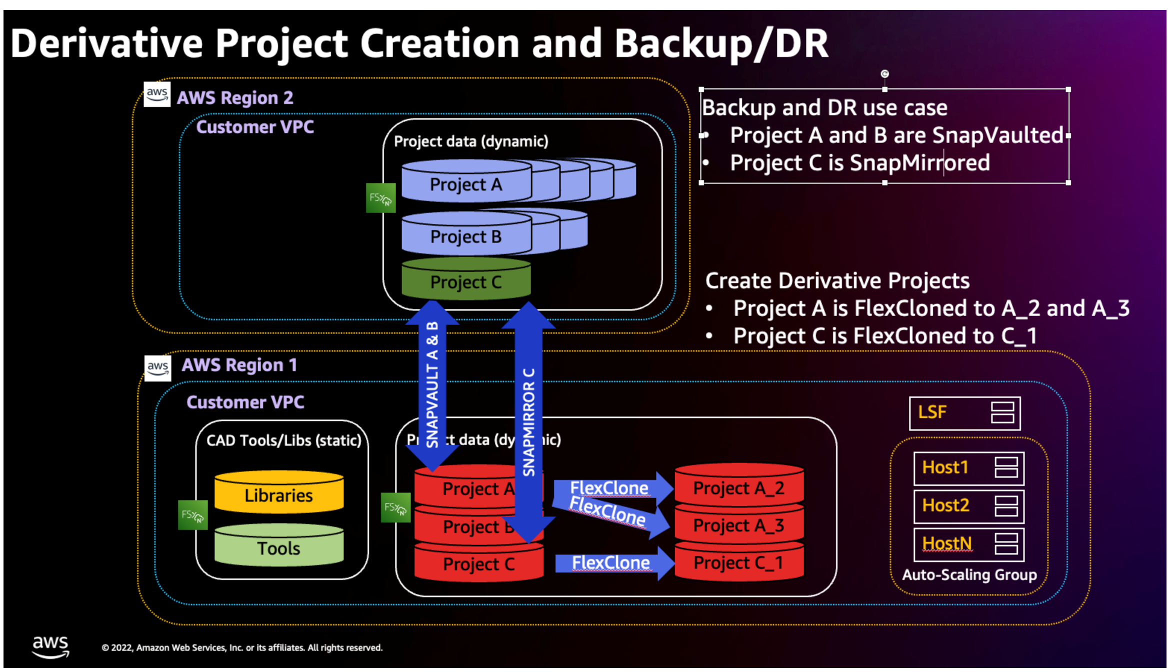


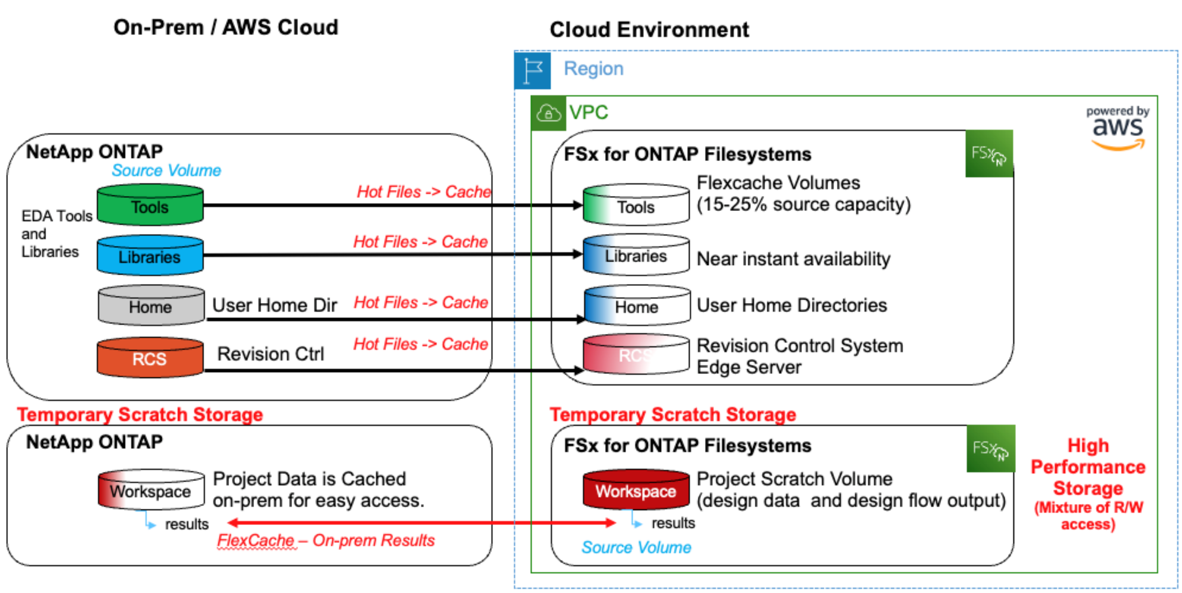

**Figure 9: Multi-site (multiple AWS regions) architecture using FSx for ONTAP.**



**Note:** A derivative project uses data from a previous project as opposed to fresh design data.

More specifically, the EDA directories (project, tools, and libraries) are replicated between either two AWS regions or on-premises and AWS as depicted in Figure 10.

**Figure 10: EDA directory replication using ONTAP technologies in a multi-site setup.**




## Multi-region data management architecture at enterprise scale

### Scaling tools and libraries

Tools and libraries are typically shared resources on-premises and therefore should also be in the cloud. Advanced semiconductor manufacturing processes (3nm, 5nm, and 7nm) increase design complexity and scale. Therefore, design technology libraries have become very large, as the complexities of representing transistor performance, power, and geometries have caused the files to grow significantly. As a result, library loading times have become a challenge.

The demand on the library volumes when tool flows start-up on 10,000 to 100,000 server cores simultaneously can create a boot storm on the storage filers, leading to high read latency. This is further compounded by sharing tool and library volumes across multiple design flows, projects, and teams. Scaling FlexCache volumes to spread the read load out across the parallel compute cores can dramatically improve library and tool load times.

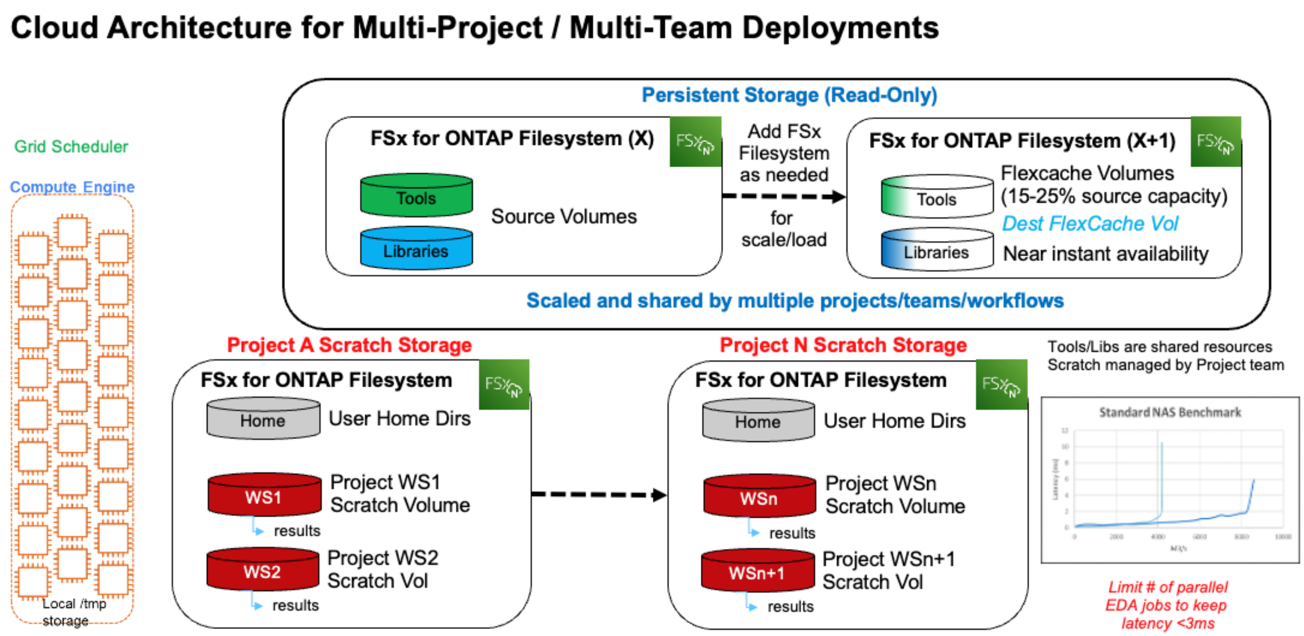
AWS recommends configuring tools and library volumes (FSx for ONTAP file systems) in a multi-AZ configuration. This configuration provides resiliency against FSx for ONTAP node and AZ failure. It is a best practice to replicate tools and libraries across multiple FSX for ONTAP file systems to achieve performance at scale and increase availability. In addition, NetApp recommends using FSx for ONTAP-native backup for tool and library volumes for improved reliability, durability, and uptime.

### Scratch volumes at scale

Scratch volumes require maximum read-write performance. It is a best practice is to use a FlexGroup volume on a single-AZ FSx for ONTAP file system. FSx for ONTAP file system performance metrics are provided here for reference.

FSx for ONTAP file systems should be deployed on a per flow, team, or project basis. Figure 11 shows how shared FSx for ONTAP file systems can be scaled horizontally by using FlexCache volumes for tools, libraries, and other read-heavy workloads. As more and more EC2 instances are elastically scaled up to meet increasing design workload requirements, the tool and library FlexCache volumes can be quickly scaled up and UNIX mount points DNS load balanced to optimize performance. As EC2 instances are scaled back down, FlexCache volumes can be scaled back as well.

**Figure 11: FSx for ONTAP architecture at scale for EDA directories.**




# Amazon FSx for NetApp ONTAP Setup

This section describes the current best practices for optimizing FSx for ONTAP for high-performance, large-scale EDA workloads.

## FSx for ONTAP file system creation setup

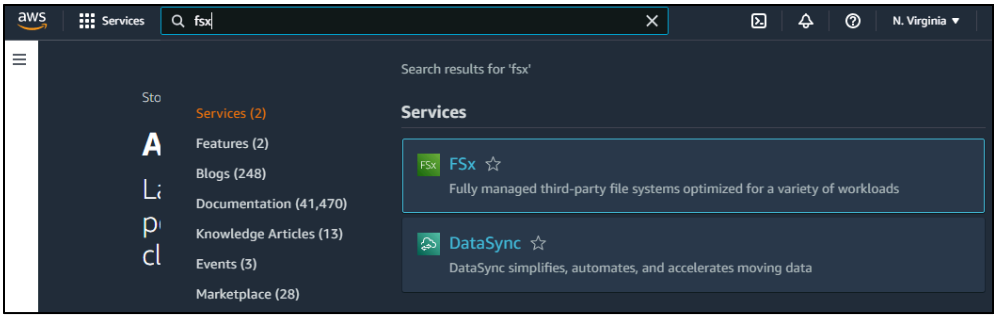
- FSx for ONTAP file system for scale-out tool and library FlexCache volumes

  This configuration is optimized for very high read-only performance as well as high availability.
- FSx for ONTAP optimized for maximum I/O for scratch volumes

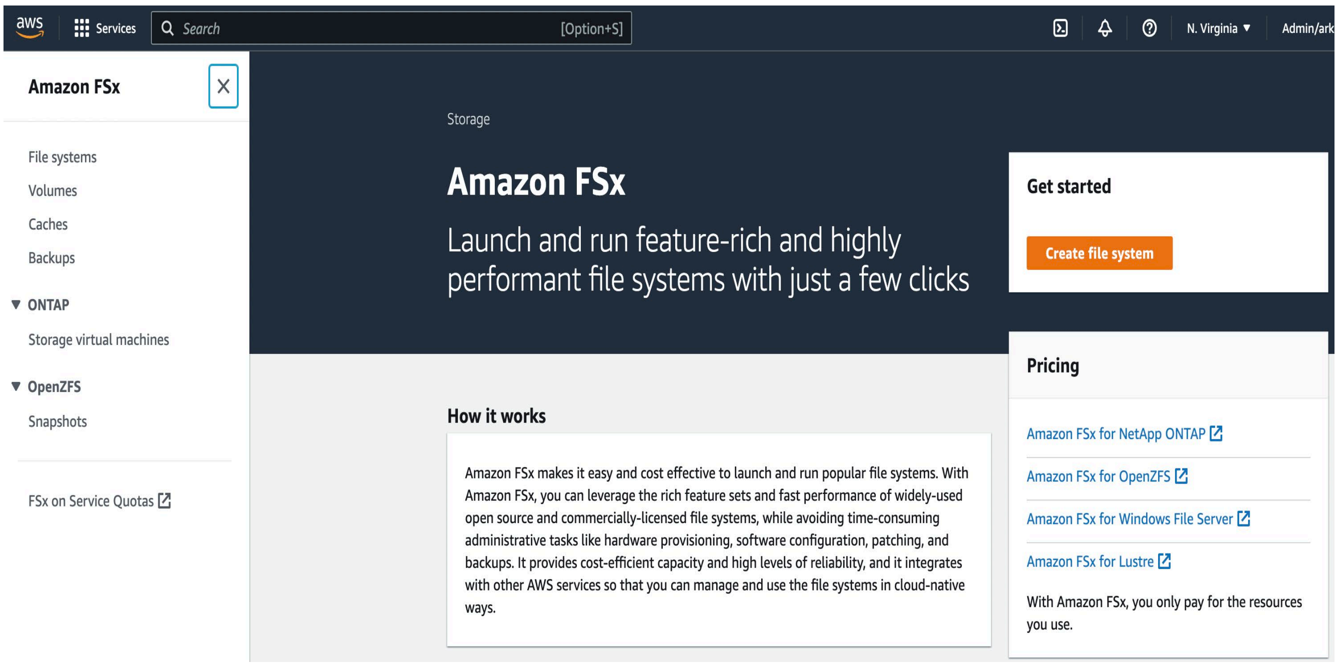
  This configuration is optimized for maximum I/O throughput for read-write performance at scale.

### Open FSx for ONTAP console

1. Open the FSx for ONTAP console by clicking here.

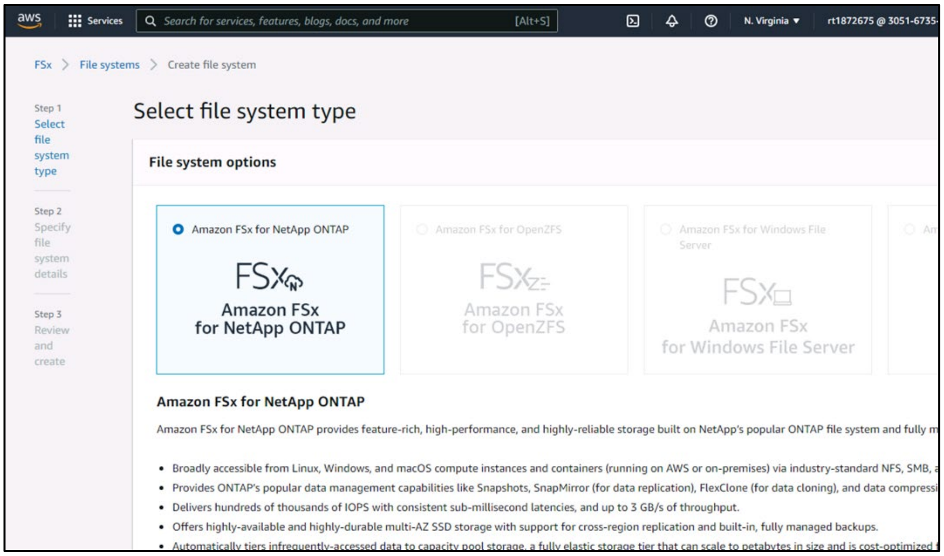



2. Click the FSx service under **Services**.

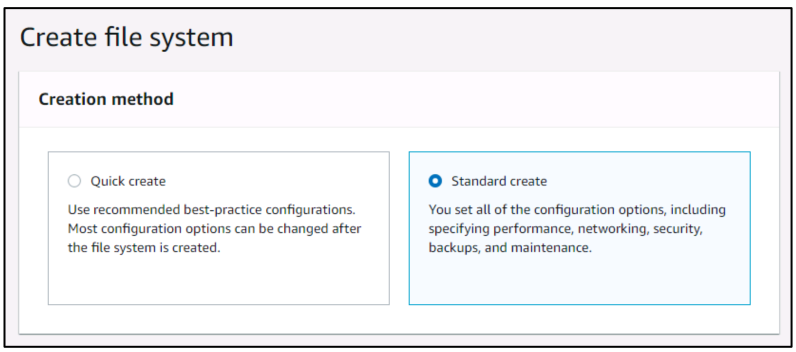



## Create a new FSx for ONTAP file system

1. Select **Create file system**, and then select **Amazon FSx for ONTAP**.



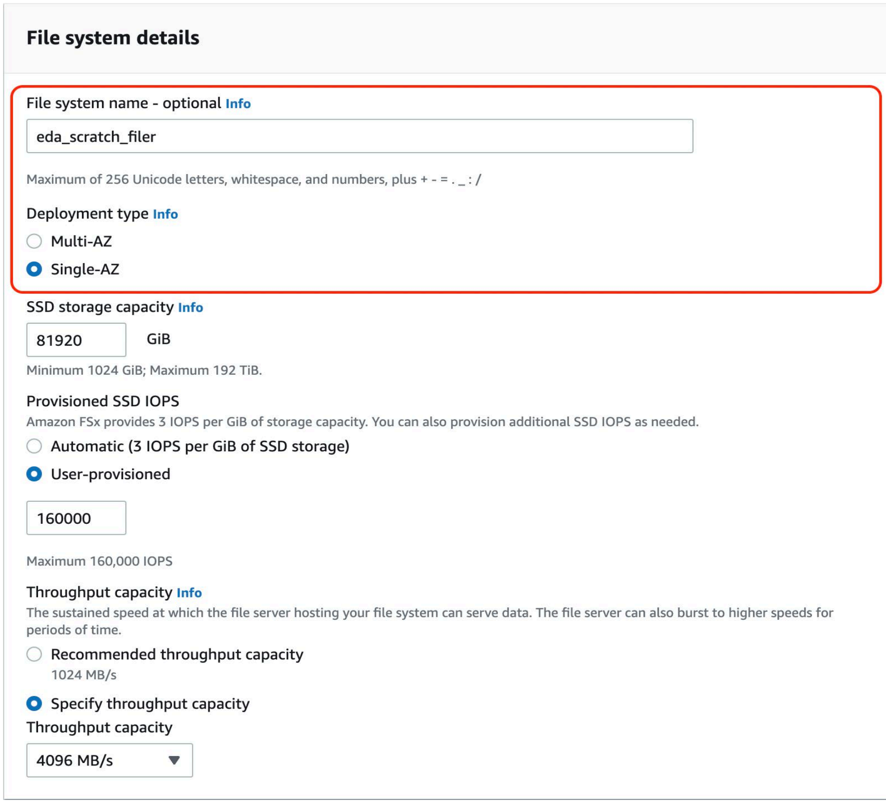

2. Under **Creation Method**, select **Standard create**.




Choosing **Standard create** provides additional customization options needed for optimal EDA performance.

## File system details

1. Specify a File system name. This setting is optional, but it is a best practice is to provide a meaningful name for the file system.




2. Select the single-AZ deployment type for the scratch volume and multi-AZ for tools, libraries, and projects.

**Note:** Both types of deployment (single AZ and multi AZ) support local NVMe for read caching. For more information, see [Amazon FSx for NetApp ONTAP performance](Amazon FSx for NetApp ONTAP performance).

## Specify capacity and storage performance requirements

For EDA workloads, it is a best practice for peak performance to start with an FSx for ONTAP file system configured for maximum IOPS and throughput performance and then to test workloads at scale. If the workloads drive higher volume latencies than desired, then you might need to load balance workloads across multiple file systems. There might be a cost optimization opportunity if file system performance is above expectations. Provisioned SSD IOPS and throughput capacity can be increased or decreased after file system creation to save storage costs.

It is typically easier to run more EDA jobs and overload the file system to find the maximum load the largest FSx for ONTAP file system can support and then to scale back the number of EDA jobs running in parallel. If the system is not sized for adequate performance, the FSx for ONTAP volume metrics in CloudWatch will show high volume latencies, which affects the total number of parallel jobs that can run before hitting the knee in the latency curve.

NetApp recommends starting at full-scale performance, because in most cases file system performance is the limiting constraint for EDA workloads.

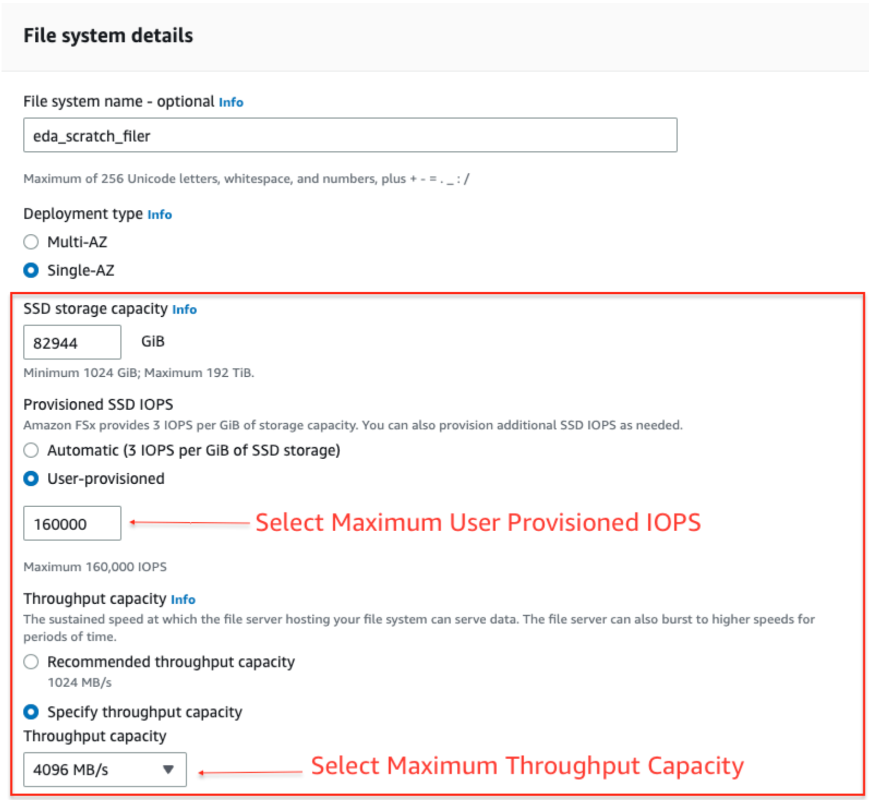
1. To provision the required SSD storage capacity, specify the capacity in GiB with a maximum of 192 TiB.

   For example, to create an 81TiB volume (81 TiB x 1024 GiB), type 82944 in the SSD storage capacity dialog box.

2. To specify provisioned SSD IOPS, select User-Provisioned IOPS, and then type 160000 into the user-provisioned textbox. The maximum allowed value is 160,000 IOPs. [Check the AWS documentation for regional availability]().

**Note:** A minimum of 5TB of space must be provided to choose 160K IOPS.

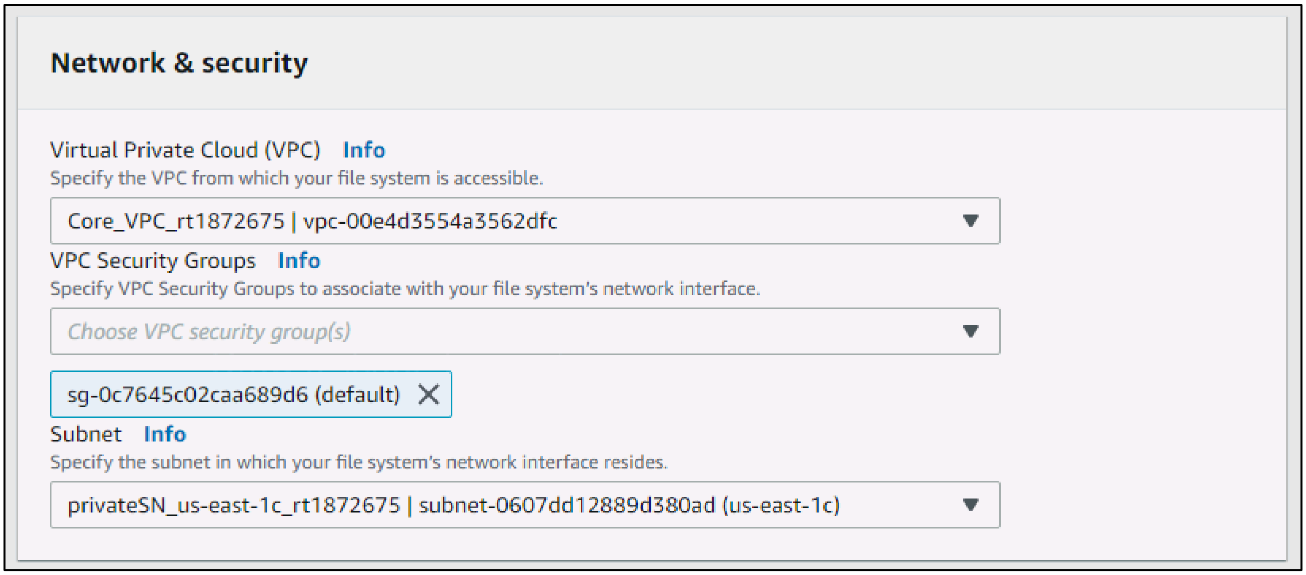
3. Select **Specify throughput capacity**, and then select **4096 MB/s** from the **Throughput capacity** pulldown. This is the maximum allowed value.




## Network and security

NetApp makes no EDA-specific best practices requirements for the network and security. All options are specific to the VPC design.

- VPC
- VPC security group. See the [documentation]() for VPC security group configuration.
- Subnet

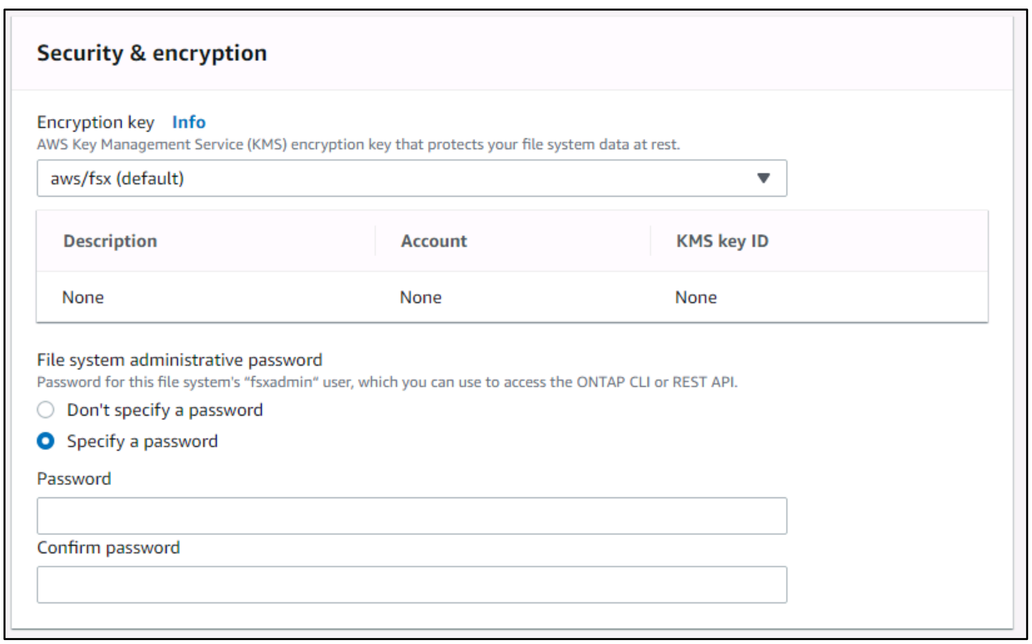



## Security and encryption

FSx for ONTAP supports nitro-based encryption in transit when accessed from nitro instances in supported regions. This encryption happens without any performance impact. Hence nitro-based instances are recommended for customers requiring encryption in transit.

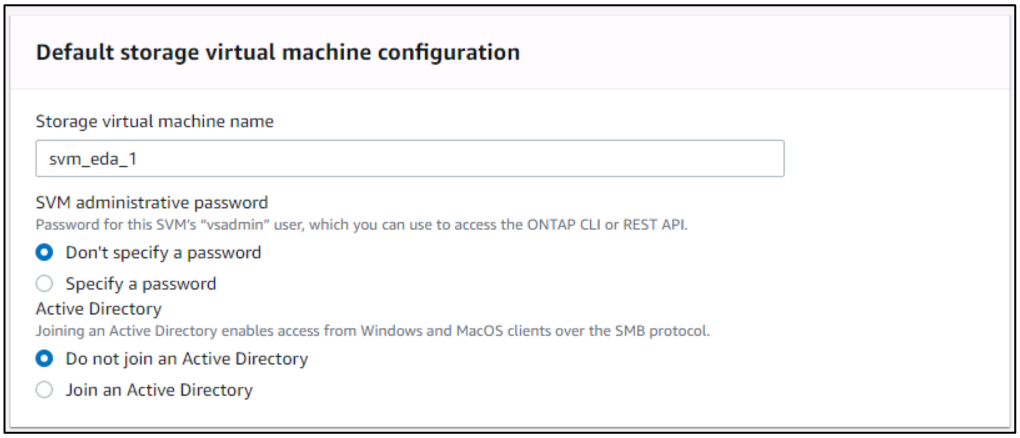
FSx for ONTAP also supports Kerberos-based encryption is supported on FSX for ONTAP with NFSv4. NFSv4 is less performant compared to NFSv3 for EDA workloads. Use of Kerberos has further performance impact.

Further information on Encryption of Data in transit is available on [FSx for ONTAP documentation.]()



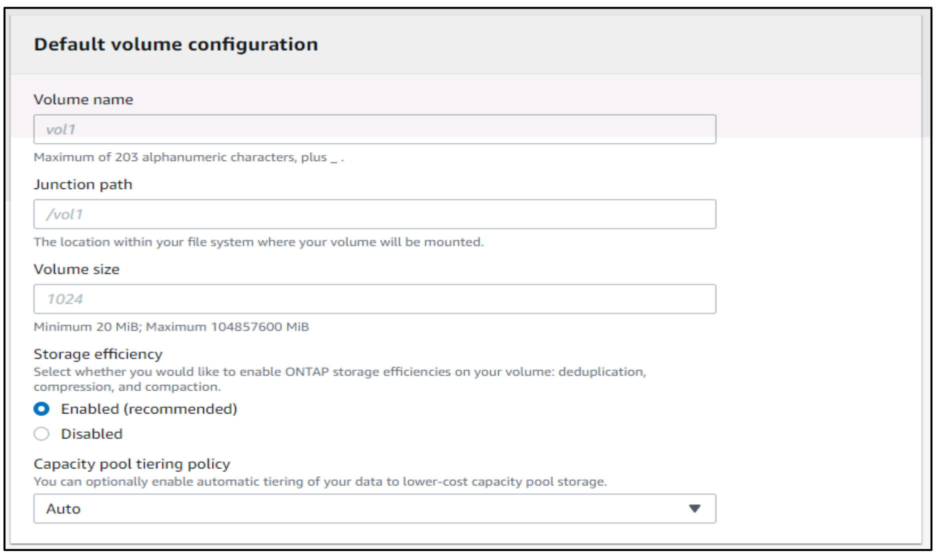
## Default SVM

1. To specify a storage virtual machine name, select the **Storage Virtual Machine Name** textbox, and then type a valid SVM name.
2. There are no EDA-specific recommendations for SVM password settings or for for Active Directory.



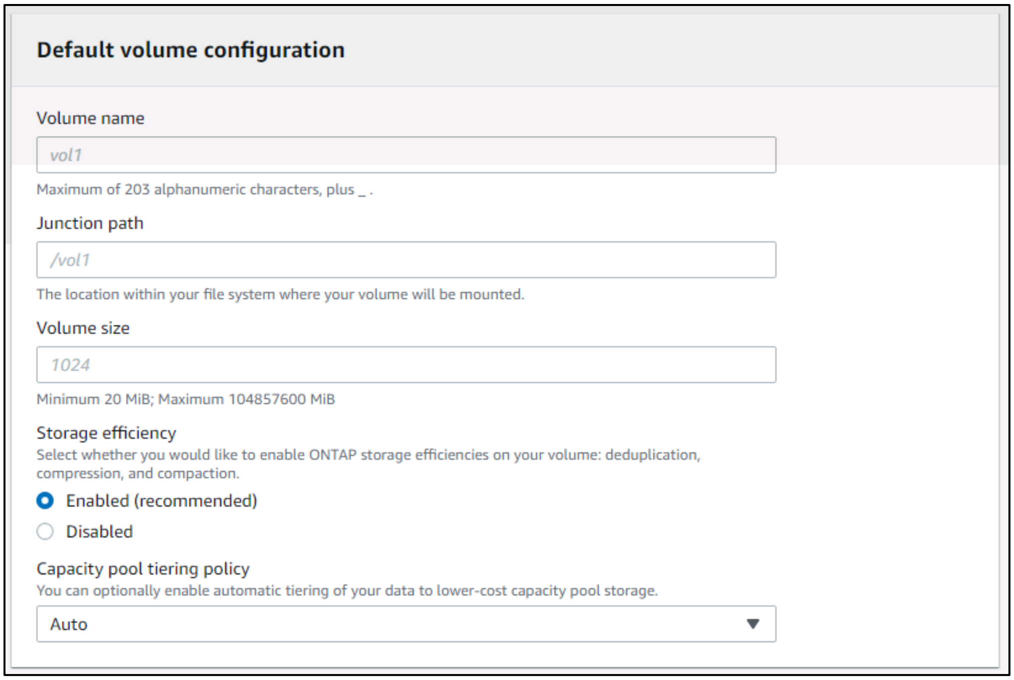

## Default volume configuration

1. For the volume configuration, fill in the desired volume name, junction path, and volume size.




2. EDA workloads benefit from enabling FSx for ONTAP in-line storage efficiencies. Compression, deduplication, and compaction lead to better storage utilization, increased operational efficiency, which results in lower storage costs. Select **Enabled** (recommended).
3. Nex set the capacity pool tiering policy. EDA workloads often contain 60% to 80% cold data. Enabling the Auto tiering policy for non-scratch volumes can help reduce storage costs. For scratch volumes, NetApp recommends setting the capacity pool tiering policy to **auto**. This enables a cost-efficient configuration. However, the capacity pool tiering policy can be set to **None**, if performance is a priority.

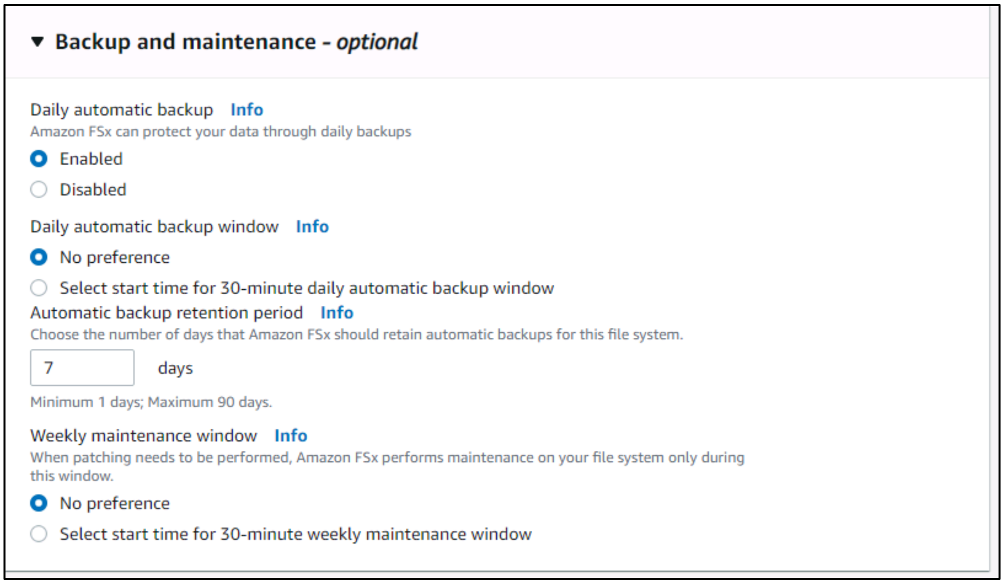
   Select **Capacity pool tiering policy**, and then select **Auto** from the pulldown.




## Backup and maintenance

Snapshot copies are enabled by default on your volumes using the default Snapshot policy. It is a best practice to review and revise your Snapshot policies when creating or updating volumes. Scratch volumes might not benefit from regular automated Snapshot copies. Scratch volumes often see a high file-change rate that can result in Snapshot copies consuming a significant amount of available storage capacity. Snapshot copies can be enabled or disabled after volume creation.

FSx for ONTAP also offers a daily automatic backup feature. Daily backups leverage Snapshot copies. The default Snapshot policy is enabled even without enabling daily backups. The FlexCache source volume should be backed up, but the FlexCache destination does not need to be because it caches the data from the source volume. Scratch volumes do not need to be backed up. Backup for other volumes (tools and libraries, projects, and home dir) is recommended.



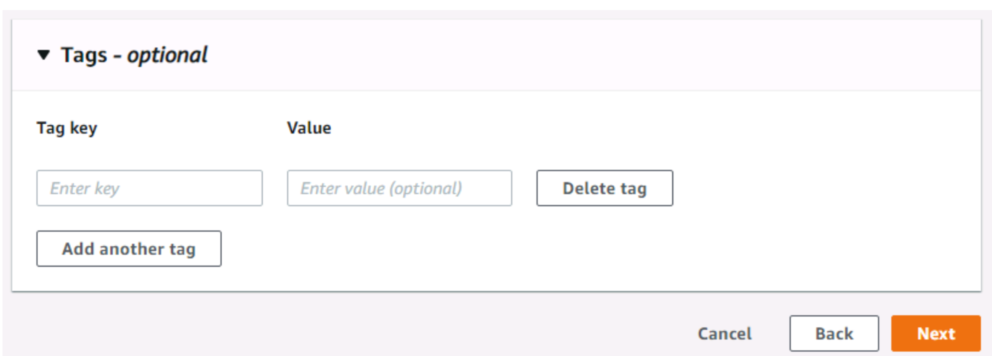

## Tags

Creating tag key-value pairs can provide searchable labels for billing, file system ownership, and so on. There are no EDA-specific recommendations for tags.

See the following example tags for an EDA environment:

**Key: Value pair**

- project_name: projecta
- eda_workflow: vcs_nightly_regressions
- workflow_owner: projecta_verification_team




## Review and approve

Review the Create File System Summary.

If the settings are correct, click **Create file system.**

**Summary**

Verify the following attributes before proceeding

| Attribute | Value | Editable after creation |
|---|---|---|
| File system type | Amazon FSx for NetApp ONTAP | ⊗ |
| File system name | eda_scratch_filer | ✓ |
| Deployment type | Single-AZ | ⊗ |
| Storage type | SSD | ⊗ |
| SSD storage capacity | 82,944 GiB | ✓ |
| Minimum SSD IOPS | 160000 IOPS | ✓ |
| Throughput capacity | 4096 MB/s | ✓ |
| Virtual Private Cloud (VPC) | vpc-0a33d2062627d42a0 | ⊗ |
| VPC Security Groups | sg-00c944a3d5e416882 | ✓ |
| Subnet | subnet-06dfde8daf842d337 | ⊗ |

**Tags**

< 1 >

| Key | Value |
|---|---|
| No data | |

**Storage virtual machines (SVMs)**

| Attribute | Value | Editable after creation |
|---|---|---|
| Name | svm_eda_1 | ✓ |
| Fully qualified domain name | | ✓ |
| DNS server IP addresses | | ✓ |
| Service account username | | ✓ |
| Delegated file systems administrators group - optional | | ✓ |

Cancel Back **Create file system**

## Post FSx for ONTAP creation setup

The following steps are run on each FSx for ONTAP file system after the file system creation process completes.

EDA workloads significantly benefit from FlexGroup volumes for maximum scale and performance. FSx-NetApp ONTAP does not yet support creation of FlexGroup or FlexCache volume types from the AWS Management Console (this may change with later FSx-NetApp ONTAP releases)

FlexGroup and FlexCache volumes must be provisioned via the FSx for ONTAP ONTAP CLI or via ONTAP REST API.

ONTAP CLI commands are shown below.

## Recommended EDA-specific options (pre-volume provision)

NetApp recommends enabling 64-bit identifiers as follows:

```
ONTAP cli> vserver nfs modify -vserver [VSERVER_NAME] -v3-64bit-identifiers enabled
```

## Create FlexCache volume for tool and library volumes

FlexCache volumes can enable almost instant replication (caching) of on-premises tool and library volumes to AWS. Each on-premises source tool or library volume requires a FlexCache volume. If there are multiple tools and library volumes on-premises, then multiple FlexCache volumes are needed in the cloud.

Setting up an ONTAP cluster and SVM peering relationship between two ONTAP file systems is a prerequisite for creating a FlexCache instance. The NetApp ONTAP documentation provides instructions for setting up peering relationships.

After the cluster and SVM peering relationships have been set up and verified, FlexCache volumes (burst to AWS or multi-region on AWS) can be created using the following command:

```
ONTAP cli> volume flexcache create -origin-vserver [ORIGIN_SVM_NAME] -origin-volume
[ORIGIN_VOLUME_NAME] \
                                   -vserver [DESTINATION_SVM_NAME] -volume
[DESTINATION_VOLUME_NAME] \
                                   -size [VOLUME_SIZE] -junction-path [JUNCTION_PATH] \
                                   -aggr-list aggr -aggr-list-multiplier 8 \
                                   -policy [EXPORT_POLICY]
```

## Set atime-update on origin or source to false (ONTAP 9.11.1 or earlier)

Set atime-updates on the origin or source volume to false. To avoid invalidations on files that are cached when there is only a read at the origin, turn off the last accessed time updates on the origin volume. If the origin or source volume is running ONTAP 9.11.1 or later, the creation of a FlexCache automatically sets atime-update to false. From the origin or source volume ONTAP CLI (ONTAP 9.11.1 or earlier), run the following command:

```
ONTAP cli> volume modify -vserver origin-svm -volume vol_fc_origin -atime-update false
```

In releases earlier than NetApp ONTAP 9.11.1, it is a best practice is to disable atime-update on the FlexCache origin volumes to reduce the amount of data being updated in the cache because of read access. Failure to disable atime-update on the source volume can lead to overly aggressive cache eviction. For example, simple reads at the source volume will evict the data from the cache.

NetApp ONTAP 9.11.1 and later provides an improved workflow that can allow for atime-updates to be propagated across the environment and not affect evictions at the cache. Even reads at the cache are able to update atime on the origin. The cache never is an authority on atime, only the origin is.

After upgrading the Origin/Source volume to ONTAP9.11.1 or later, if there is no use for atime, then leave atime-update off.  If there are workflow decisions made based on atime, then use the knobs created in 9.11.1 appropriately for your need.

## Create FlexGroup volume for read-write heavy workload

Create one or more FlexGroup volumes for scratch. If required, create volumes for tools, libraries, and the project as well.

```
FSx for ONTAP file system
ONTAP cli> volume create -vserver [SVM_NAME] -volume [VOLUME_NAME] \
                         -size [VOLUME_SIZE] -junction-path [JUNCTION_PATH] \
                         -aggr-list aggr1 -aggr-list-multiplier 8 \
```

```
                        -policy [EXPORT_POLICY] \
                        -files-set-maximum true [-supporttiering true]
```

More information on FlexGroup volume creation can be found here.

FlexGroup volumes do not currently support backups. If backups are needed, you should consider using SnapVault to another FSx for ONTAP file system.

## Large file considerations

The aggr-list-multiplier specifies the number of constituent volumes per node or aggregate pair. Eight constituent volumes are recommended for EDA workloads. If an 80TB volume is provisioned, each constituent will be 10TB in size. A large file is any file that is 1% to 5% of the member constituent volume. In this example, a file large is any file larger than (80TB/8 x 5%) or 0.5TB (500GB). In general, it is better to provision larger thin-provisioned FlexGroup volumes than smaller volumes.

## Increase iNode count

EDA environments can have millions of files, requiring an increase in the inode value to the maximum in many cases. It is best practice to constantly monitor the inode utilization using Amazon CloudWatch metrics. FilesUsed Volume metrics can also be used for the purpose. NetApp recommends increasing the inode count when 80-90% utilization is observed.

Inode defaults and maximums according to FlexVol size.

**Table 2: Inode defaults and maximums according to FlexVol size.**

| FlexVol volume size | Default inode count | Maximum inode count |
|---|---|---|
| **20MB*** | 566 | 4,855 |
| **1GB*** | 31,122 | 249,030 |
| **100GB*** | 3,112,959 | 24,903,679 |
| **1TB** | 21,251,126 | 255,013,682 |
| **7.8GB** | 21,251,126 | 2,040,109,451 |
| **100TB** | 21,251,126 | 2,040,109,451 |

- FlexGroup member volumes should not be any smaller than 100GB in size.

## Mount volumes

NFSv3 continues to be the standard for EDA workloads because it is more performant than NFSv4.x. The following sections provide mount commands for read-heavy tool and library volumes and for providing the fastest possible performance for read-write scratch volumes.

See the section “Further Readings” for technical reports (TRs) for NFS and FlexGroup Best Practices for further guidance for optimizing NFS mounts for EDA and HPC workloads.

## Mount command for read-optimized tool and library volumes

To improve file and metadata loading times as well as to decrease the load on the ONTAP storage system, NetApp recommends using aggressive compute server caching for tool and library volumes that are mostly read-only in nature. The NFS mount options `nocto, actimeo=600` have been shown to reduce metadata I/O for EDA workloads by as much as 90%. These options can also dramatically improve the load time for tools and libraries, which is becoming an increasingly big challenge as semiconductor technology library sizes grow dramatically in number and size at 3, 5, and 7nm design nodes.

EDA workflows often make heavy use of Makefile or Makefile-like dependencies tracking. The metadata I/O is 60-80% of overall IOPS. In addition to Makefile-like dependency checks, EDA and scripting tools like Perl and Python often scan long files or directory search paths in the modules to load. The result is additional metadata I/O transactions used in the search for files and directories. This can put an unnecessary load on the file system.

Aggressive NFS client-side caching works based on the assumption that tool and library volumes are static or unchanging. New tools and libraries are typically installed in directories next to older versions and are seldom over-written or modified and it is considered poor practice for EDA workloads to pull in the latest libraries mid analysis. EDA flows typically specify a set of tools and libraries for a given job, and then those versions are used throughout the run.

For more information, see this blog article: What does actimeo mean during NFS mount in Linux?

NetApp recommends the following mount command to maximize server-side caching using the nocto and actimeo options.

Review and apply this best practice per the EDA flow's requirements. Overly aggressive server-side caching can lead to slow cache-updating issues.

From the client side, run the following command:

```
%> mount -t nfs \
  -o "nocto,actimeo=600,hard,rsize=262144,wsize=262144,vers=3,tcp,mountproto=tcp" \
         <ontap svm>:/<volname> /mnt/<volname>
```

**Note:** See EDA – Best Practices in ONTAP for more information on command-line values.

## Mount command for mixed read-write optimized performance on scratch volumes

The following mount option is optimized for high-performance NFS read-write transactions. Unlike read-only tool and library volumes, aggressive file caching is typically not recommended.

The following NFS v3 mount command is recommended for EDA scratch volumes workloads:

```
%> mount -t nfs \
         -o "hard,rsize=262144,wsize=262144,vers=3,tcp,mountproto=tcp" \
         <ontap svm>:/<volname> /mnt/<volname>
```

**Note:** See EDA – Best Practices in ONTAP for more information on command-line values.

NFS v4.x is not recommended for EDA workloads and may result in as much as 20% performance penalty.

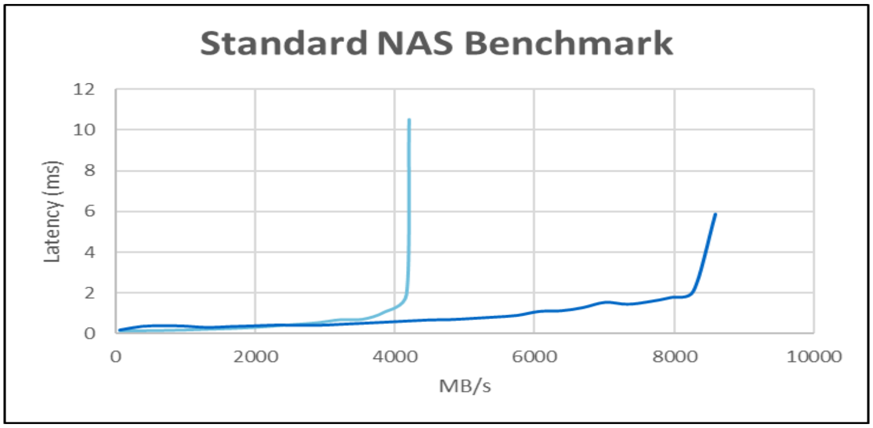
## Performance monitoring

Performance monitoring is critical for EDA workloads. The single most important metric is volume read-write latency. If you see that the latency of the volume is not at an acceptable level for the EDA application, it means the volumes are under heavy I/O workload. Therefore, you should scale-up and distribute the workload across multiple FSx for ONTAP file systems.

High latency conditions can occur to high I/O loads from EDA jobs running in parallel. NetApp recommends testing EDA workloads running on FSx for ONTAP at scale to determine the I/O limits. Figure 12 depicts an NAS workload that has been pushed to its performance limits. The knee of the performance curve is the point at which the controller is operating outside of its performance profile and latencies start to increase rapidly. Operating in this state for long periods of time affects the EDA job runtime performance (wall clock).

**Note:** Figure 12 is a generic NAS performance profile graph for illustration purposes only and does not represent actual FSx-NetApp ONTAP performance data.

**Figure 12: Standard NAS benchmark.**




## Quick latency monitoring with ONTAP CLI

There are multiple ways to measure real-time ONTAP performance. One way is to simply report latency using the ONTAP CLI. The following `qos statistics volume latency show` command runs until Ctrl-C is pressed to stop reporting and then outputs a report as follows:

```
cluster1::> qos statistics volume latency show -iterations 100 -rows 3
Workload            ID  Latency    Network  Cluster       Data      Disk  Qos Max    Qos Min      NVRAM      Cloud  FlexCache    SM Sync         VA
--------------- ------ --------   -------- --------  -------- -------- -------- ---------- ---------- ---------- ---------- ---------- ----------
vol2               999  44.24ms    44.05ms      0ms  190.00us      0ms      0ms        0ms        0ms        0ms        0ms        0ms        0ms
-total-              -  38.89ms    38.63ms      0ms  256.00us      0ms      0ms        0ms        0ms        0ms        0ms        0ms        0ms
vol2               999  64.47ms    64.20ms      0ms  266.00us      0ms      0ms        0ms        0ms        0ms        0ms        0ms        0ms
vol1              1234  27.28ms    27.03ms      0ms  253.00us      0ms      0ms        0ms        0ms        0ms        0ms        0ms        0ms
vslvol0            111  23.72ms    23.47ms      0ms  249.00us      0ms      0ms        0ms        0ms        0ms        0ms        0ms        0ms
```

More detail on the command can be found here.

Start the command either before or during EDA job submission to check the real-time volume latency. Extended periods of latency greater than the acceptable latency could be a sign of a heavily loaded storage file system.

## Monitoring and reporting

You can use the following services and tools to monitor Amazon FSx for ONTAP usage and activity:

- **Amazon CloudWatch.** You can monitor file systems using Amazon CloudWatch, which automatically collects and processes raw data from FSx for ONTAP into readable metrics. These statistics are retained for a period of 15 months so that you can access historical information and see how your file system is performing. You can also set alarms based on your metrics over a specified time period and perform one or more actions based on the value of the metrics relative to thresholds that you specify. For more details, see the [FSX for ONTAP monitoring documentation.](#)

- **NetApp Cloud Insights.** You can monitor configuration, capacity, and performance metrics for your FSx for ONTAP file systems using the NetApp Cloud Insights service. You can also create alerts based on metric conditions.
- **NetApp Harvest and NetApp Grafana.** You can monitor your FSx for ONTAP file system by using NetApp Harvest and NetApp Grafana. NetApp Harvest monitors ONTAP file systems by collecting performance, capacity, and hardware metrics from FSx for ONTAP file systems. Grafana provides a dashboard where the collected Harvest metrics can be displayed.
- **AWS CloudTrail.** You can use AWS CloudTrail to capture all API calls for Amazon FSx as events. These events provide a record of actions taken by a user, role, or AWS service in Amazon FSx.

### Capacity reporting

Reporting on the size and utilization of a FlexCache volume can be achieved using the command `show –fields <fields to report>` from the ONTAP CLI. There are RESTful equivalents as well.

Here is an example with a few fields populated. In the following example, the AWS FSx for ONTAP `proj` volume is a FlexCache instance of the on-premises `proj` source volume.

```
ONTAP: :> volume show -volume <volname> -fields size,used,available,percent-used,files-used,files
```

In the following example, the on-premises `proj` volume is a FlexGroup volume and the AWS `proj` is a FlexCache volume. After reading (or pre-warming) some of the files in the FlexCache, you can see the source volume capacity used and the much smaller FlexCache capacity used.

```
ONTAP CLI on-Prem::> volume show -volume proj -fields size,used,available,percent-used,files-
used,files
vserver    volume size available used    percent-used files    files-used
---------- ------ ---- --------- ------- ------------ -------- ----------
svm_onPrem proj   1TB  46.47GB   15.85GB 1%           31876696 226372

ONTAP CLI FSx-N::> volume show -volume proj -fields size,used,available,percent-used,files-
used,files
vserver    volume size  available used   percent-used files   files-used
---------- ------ ----- --------- ------ ------------ ------- ----------
svm_awsFSx-N proj   150GB 147.0GB   2.96GB 1%           4669368 792
```

# Where to find additional resources

To learn more about the information that is described in this document, review the following documents and/or websites:

- FSx for ONTAP Documentation on-line

  https://docs.aws.amazon.com/fsx/latest/ONTAPGuide/what-is-fsx-ontap.html
- Amazon FSx for NetApp ONTAP performance

  https://docs.aws.amazon.com/fsx/latest/ONTAPGuide/performance.html
- Electronic Design Automation best practices

  https://www.netapp.com/media/19368-tr-4617.pdf
- NetApp ONTAP FlexGroup volumes best practices and implementation guide

  https://www.netapp.com/pdf.html?item=/media/12385-tr4571pdf.pdf
- NFS in NetApp ONTAP Best practice and implementation guide

  https://www.netapp.com/pdf.html?item=/media/10720-tr-4067.pdf

- Amazon FSx for Netapp ONTAP Monitoring Amazon FSx for ONTAP https://docs.aws.amazon.com/fsx/latest/ONTAPGuide/monitoring_overview.html

# Document history

| Date | Description |
|---|---|
| **Jan, 2023** | Brief description of revisions. |
| **Feb, 2023** | First publication |