Making Everything Easier!™

Avere Systems Special Edition

Cloud-Enabled Data Center

FOR DUMMIES®

A Wiley Brand

Learn to:

- Converge local storage and compute infrastructures with the cloud
- Adopt cloud technologies at a pace that makes sense for your business
- Get up to speed on NAS, hybrid cloud, and cloud bursting

Brought to you by

AVERE

About Avere

Avere is radically changing the economics of data storage.

Avere's hybrid cloud-enablement solutions give companies the ability to end the rising costs and complexity of scaling data storage and compute via the freedom to store and access files anywhere—in the cloud or on premises—without sacrificing the performance, availability, or security of enterprise data.

Based in Pittsburgh, PA, Avere is led by veterans and thought leaders in the data storage industry.

Cloud-Enabled Data Center

FOR DUMMIES®
A Wiley Brand

Avere Systems Special Edition

by Sharif Nijim

FOR DUMMIES®
A Wiley Brand

Cloud-Enabled Data Center For Dummies®, Avere Systems Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

ISBN 978-1-119-36136-7 (pbk); ISBN 978-1-119-36137-4 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Publisher's Acknowledgments

We're proud of this book and of the people who worked on it. Some of the people who helped bring this book to market include the following:

Project Editor: Martin V. Minner
Senior Acquisitions Editor: Amy Fandrei
Editorial Manager: Rev Mengle
Business Development Representative: Susan Blessing
Production Editor: Antony Sami
Avere Systems Contributors: Joy Burd and Jeff Tabor

Table of Contents

Introduction

Traditional companies have established an abundance of expertise developing applications and operating local data centers. Over the past decade, two factors have been challenging IT groups. The first is the incredible rate at which data is growing. The second is the rise of the public cloud, which represents the biggest shift in computing of the last decade.

The ability to have fast access to data stored in on-premises network attached storage (NAS), object storage, or in the public cloud is critical to achieving strategic goals. With increasing competition putting pressure on cycle times, companies need innovative ways to bring compute and storage assets to bear without expanding or relying on data centers. By leveraging the advantages of cloud computing, companies can increase storage and compute capacity while saving money. If attaching to the public cloud while maintaining on-premises storage performance sounds intriguing, this book is for you.

About This Book

Cloud-Enabled Data Center For Dummies, Avere Systems Special Edition, gets you up to speed on NAS, hybrid cloud, and cloud bursting. Discover how you can incorporate cloud computing and storage technologies into your local data center while retaining a familiar look and feel for your staff. With Avere Hybrid Cloud NAS, you can converge your local storage and compute infrastructures with the cloud. As on-premises equipment blends with what's available in the cloud, hybrid cloud helps you optimize your operations. You can adopt cloud technologies at a pace that makes sense for your business.

Icons Used in This Book

As you go through this book, keep a lookout for the following icons in the margins. They denote knowledge nuggets to help fire your imagination.

This icon denotes useful bits of information to squirrel away in the back of your mind. You'll also find a few factoids sprinkled in that you can use to wow your colleagues.

This identifies a taste of the technical. Not enough detail to be a hybrid cloud engineer; just enough detail to help you ask the right questions.

This icon points out things to consider as you develop your future infrastructure plans.

This icon highlights pitfalls to watch out for — history is filled with lessons you don't want to relearn.

Beyond the Book

Avere Hybrid Cloud NAS is reinventing the way you store and access critical business data. Check out `www.averesystems.com` for further information on Avere Hybrid Cloud NAS.

Where to Go from Here

If you've never heard of NAS or hybrid cloud computing, it's a good idea to take a look at Chapters 1 and 2 first to get an introduction. Otherwise, feel free to skim the book and dive into whatever chapter interests you most. Beyond that, buckle up, sit back, relax, and enjoy the read.

Chapter 1

Understanding NAS

In This Chapter

- Reviewing network attached storage
- Managing data growth
- Sharpening the blade
- Looking for ways to resolve overloaded NAS systems

Network attached storage (NAS) has been around for decades. To be productive, humans need to share information. Early system architects had this need and built a network among their computers to share data. Today's NAS systems embody the same ageless need to share and collaborate.

Once these architects sent some bits and bytes from one computer to another, they decided it would be great for that data to stick around for a long time. That idea led to the remote storage of data, separate from the computer that actually created it. It also led to the accumulation of data that didn't need to be stored for long periods of time. People like to play safe and keep data around, "just in case." As more people hoarded more data, the problem of data management was born.

This chapter reviews what NAS is, gives you some insight into how managing data can be difficult, and offers some ideas and guidelines for managing vast amounts of data.

Getting to the Basics of NAS Storage

NAS is a method of presenting a shared area where many different computers can store their files via a network. Even the most primitive NAS system has these three basic components:

- A network, which links people's computers to the NAS.
- Physical storage devices. Traditionally, these were spinning drives. Over the past decade, modern NAS has incorporated faster solid-state drives (SSDs) with no moving parts. These drives are located inside a server, known as a *filer*.
- A person's computer attached to the storage via the network. This person can see files and folders, just like on a local hard drive. A difference is that a NAS-presented folder may also contain other people's files.

Figure 1-1 shows the basic idea behind NAS, with multiple computers connected to remote storage:

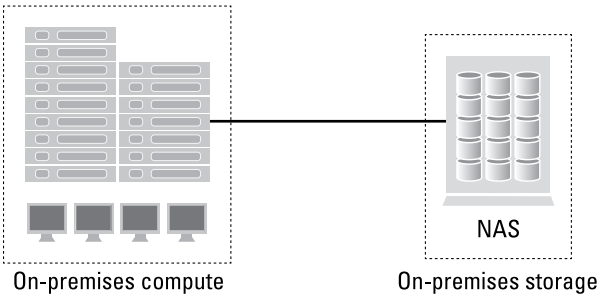


Figure 1-1: Computers talking to remote storage.

One nice thing about NAS is that tens of thousands of clients can be attached to the shared storage areas.

The ability to share data easily is fundamental to collaboration and efficiency in a workforce.

What made NAS grow so quickly is that the basic *protocol* — or language — that it uses to communicate isn't proprietary.

Because the protocol isn't proprietary, you don't have to have a particular operating system to make it work. The NAS protocol comes in two types, and both allow simultaneous, shared access to files across a network from different client computers:

- **Network File System (NFS):** This protocol was adapted as a standard for Unix-based operating systems.
- **Server Message Block (SMB):** This came about later, when Microsoft created Windows. Instead of using NFS, the company decided to extend SMB functionality with network file-serving capabilities.

Both of these protocols are widely supported and formally standardized across most operating systems and NAS vendors. A standardized protocol lets people choose and share information without worrying about compatibility issues with Linux, Windows, or Mac-based client operating systems. People are free to choose the best hardware and software for their needs.

Handling the Incredible Growth and Management of Data

Data comes in two basic types: structured and unstructured. *Structured data* has a defined format. Each record, field, and data type expressed within a structured data file follows a predefined layout. You can jump to any spot inside a structured data file. Using the predefined layout as a map, you know what the bytes you're looking at actually represent.

Unstructured data has no practical internal definition. Adding or subtracting data from a file that is unstructured is much more difficult than in a structured file. You can do it, but it isn't as easy as simply running a database query to add or delete certain records. Documents, spreadsheets, and media files are all unstructured. The application that creates the file knows how to deal with it, but in terms of the bulk management of unstructured data, you are at the mercy of the sheer number of applications that can deal with these files correctly.

As the calendar rolled to 2000, the types of data created and the size and amount of data that companies have kept on hand have skyrocketed. Leading the charge is the explosion of unstructured data (see Figure 1-2).

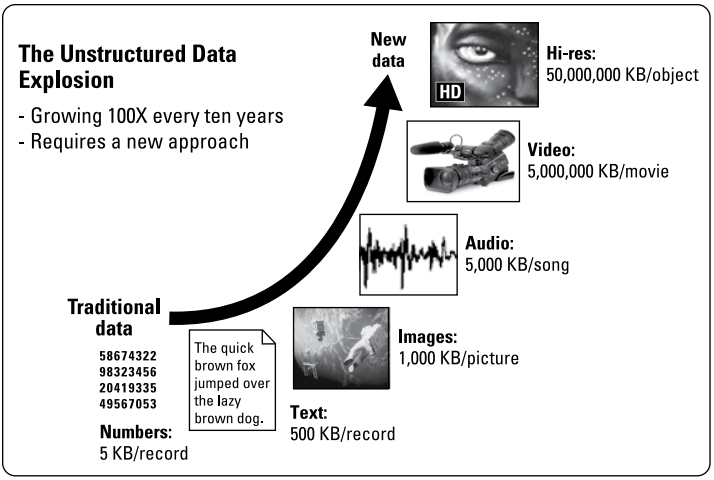


Figure 1-2: Unstructured data is growing more plentiful and bulky to deal with.

Every person with a smartphone can be viewed as an unstructured data generator — all those photos have to get stored somewhere!

You may have had to help accommodate the surge in data volume by building more and more NAS systems to hold the data. All these "NAS islands" need to keep up with both physical storage and client access. Apart from customer complaints, how do you monitor your storage efficiency and limitations? This is a common challenge felt across a variety of industries including financial services, life sciences, media and entertainment, and government.

Optimizing NAS

Disk technology has come a long way since NAS was born. Traditionally, performance gains were driven by increasing the rotational speeds of spinning disk drives. That all changed when blazing fast SSDs became commercially viable.

Companies created all-SSD NAS devices to deliver the fastest possible storage performance. Companies now offer hybrid arrays with a tier of fast SSD drives for data in active use (*hot data*). As data cools, it migrates its way to the slower, spinning disk.

Many options are available for optimizing NAS. However, if you focus only on NAS improvements, you are not taking advantage of the biggest change in IT — the rise of the public cloud.

Avere provides a scalable, all-SSD product called the FXT Series that optimizes disk-based NAS infrastructures by creating a high-performance tier. Using advanced algorithms, the FXT Series automatically moves active data to its internal SSDs to boost the performance of the applications accessing NAS, as shown in Figure 1-3. As later chapters show, the FXT Series also provides a gateway to the public cloud.

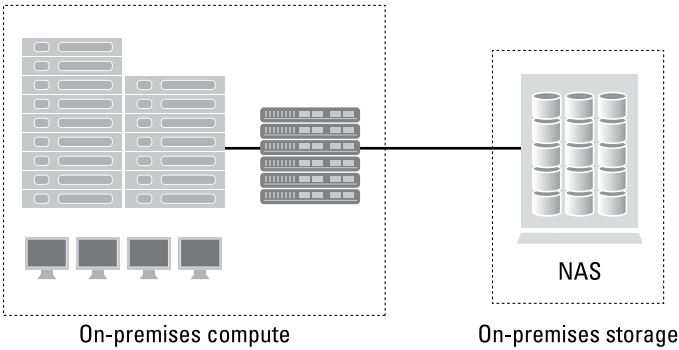


Figure 1-3: Computers talking to remote storage with Avere providing NAS optimization.

Finding an Efficient Solution to Overloaded NAS Systems

When you have all that data to manage, you have to manage it wisely. Data that was important yesterday may not be as valuable today — or it may be even more important now than it was then. You never know how business needs will change the value of your data, so dynamic data management

is crucial to adapting to changes. NAS storage has a great advantage when compared to traditional block-based storage; it is more mobile, and therefore can provide intelligent data placement. The network that NAS storage is attached to gives you the flexibility to move your files around to the location that suits them best.

Imagine having a layer of technology above all your NAS solutions that recognizes access patterns to files, then redistributes files according to changing business needs regardless of where the storage is geographically. That would be an incredibly convenient tool to have. Rest easy; this magical tool exists, so dive into the next chapter to learn all about it.

Chapter 2

Understanding Hybrid Cloud

In This Chapter

- Getting your head around the cloud
- Discussing the cloud's pluses
- Understanding that the cloud isn't a perfect world
- Looking at the differences between legacy computing and cloud computing
- Talking about cloud vocabulary

The term "cloud" came about from IT folks doing diagrams of servers and networks to show how everything is connected. When people didn't know what to draw for a certain section, they simply used a cloud-shaped squiggle. When the Internet came into the picture, IT people didn't know how to represent "the Internet," so they drew another bubble to show "We're connected to something inside this cloud." The cloud is really nothing more than a metaphor for the Internet. Most folks don't know what equipment or infrastructure they're talking to or where it lives, but everyone knows it's there.

In this book, when I refer to "the cloud," I am referring to Infrastructure as a Service (IaaS) — specifically compute, storage, and networking resources that can be used as a utility in a pay-per-use model. This chapter explores how big the cloud is, explores organizational opportunities, and discusses the similarities and differences between cloud and traditional server, storage, and network infrastructures. Touching on cost saving opportunities, I also talk about how you can use the cloud in concert with your current computing and storage platforms securely and easily.

Exploring the Size of the Cloud

The three biggest cloud providers are Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform. Much has been written about the size of these vendors, and all of them are currently enjoying broad and rapid adoption. AWS, Microsoft Azure, and Google Cloud Platform are all experiencing double digit year-over-year growth.

At its current growth rate, the cloud market is on track to eclipse the market of traditional network attached storage (NAS) providers like EMC and NetApp. What's more interesting than current adoption rates or market size is the total available market (TAM) for cloud services. Analyst estimates place the TAM for cloud services in excess of $50 billion within the next several years.

Appreciating the Advantages

A main reason major companies are adopting the cloud is that hardware doesn't reside at a company's physical site. Companies are leaving behind the headache of equipment life cycle management. With less equipment to manage, companies aren't building new data centers, buying gear to fill them, and maintaining that gear. It's like renting a car: Car rental agencies buy, store, and maintain cars. As a customer, you pay to use a car only when you need it, and you don't even have to worry about changing the oil. Thanks to the rise of Zipcar and Uber, as well as the traditional rental agencies, many people are forgoing the burden of car ownership.

Like car rental agencies, cloud providers offer services on a pay-per-use or subscription model. Cloud providers build virtual servers, and you rent them. When you're finished, you simply turn the virtual servers off. It's that simple.

Cloud services have another big benefit: cost savings. Cloud providers enjoy massive economies of scale. That's why the price you pay to rent servers or storage is incredibly cheap compared to what it would cost for you to build, power, cool, and maintain your own infrastructure. Because IT is typically a big cost center, this benefit can be significant.

The cloud allows you to provision servers and storage when you need it. If your workload peaks, you can simply rent additional capacity. Workload optimization is a big factor in an organization's cloud spend.

Pondering the Challenges

The cloud clearly has major advantages. That said, certain aspects of cloud computing make some people feel skittish. The cloud is outside your normal data center walls, which means that immediate physical security and long-term data storage security are now in someone else's domain.

When someone else physically controls your storage and compute farm, you may be nervous at first. Many cloud providers address that concern by having no single point of failure, physically or geographically. That means your applications living in the cloud are capable of running anywhere your provider operates a data center. Cloud providers have multiple physical data centers that back each other up, and redundant physical equipment is normal.

Getting your data to the cloud can be a challenge because the cloud is far away. In networking terms, "latency" (or delay) exists between your data center and the cloud. This latency can be as much as $\frac{1}{10}$ of a second. This doesn't sound like a lot, but if you are processing data in the compute cloud and need to access files in your data center, network delays of $\frac{1}{10}$ second per read really add up.

Cloud providers have great security, employing many more security experts than most enterprises. However, securing data *within* the cloud with encryption tools is up to you. Products like Avere FXT Edge filer provide the highest level of additional security by supporting AES-256 encryption and Key Management Interoperability Protocol (KMIP) to manage the encryption keys on your premises.

Framing Opportunities

From the consumer's standpoint, everything inside the cloud — networks, storage, and servers — is virtual.

Consumers never touch physical equipment, because it's defined as a piece of software. In order to maximize the cloud's true potential, your organization must nurture its ability to software-define the infrastructure that best fits your needs with customized software scripts.

The only way to manage a thousand servers as easily as ten is through automation. Having folks on staff who are adept at scripting and automation is key to a successful cloud journey. Creating scripts instead of using screwdrivers has major implications as you consider your workforce, especially in the context of hiring. You will want to grow your automation capabilities.

People adapt to new tools more easily if the tools look and feel familiar. Filer products like those offered by Avere help ease the transition for storage administrators.

Embracing cloud infrastructure means you'll have fewer screwdrivers to turn, but more software to manage. As a leader, you need to prepare your organization for this new reality. You need to take on the challenge and find ways to help your existing staff evolve.

Absorbing Cloud Storage Basics

Cloud computing resources are basically the same as those in your typical data center:

- Servers that your applications run on
- Storage devices where your data resides
- Networking resources that connect the servers, storage, and people who exist outside the cloud

As you get over "acronym overload," you'll be far more comfortable with using the services that the cloud provides. Operating cloud servers and networking is similar to what you're doing now. The major differences with the cloud are the virtually limitless compute available and how you store data. Not only are these differences cost effective; they give you ultimate flexibility.

The cloud primarily deals with object storage. The reason I say *objects* instead of *files* or *data* is because the concept of an *object* is the atomic structure that you need to fully understand before I can introduce any other details of the cloud. But, before I go into what objects are, you need to understand a couple of other terms: *namespace* and *buckets*.

Namespace

A *namespace* is a way to organizationally manage many objects within a big container. Namespaces can span physical locations and systems. A namespace contains uniquely named items, such as buckets. Namespaces help to fence off sets of unique objects from those found in other namespaces.

A namespace is like a universe. A universe knows what's inside it, but it doesn't know what's in a different universe.

Buckets

Buckets are located within a namespace. A bucket is completely unique within a namespace.

Buckets contain objects, which are like files. To refer to an object, you need the namespace, bucket, and object name, which combine to be functionally equivalent to a directory-path/filename combination.

Objects

Objects are the stored data. Objects are addressed by something called a *key-value pair,* which has two halves. The first half is essentially a filename, called the *key-name*. It exists only once within a globally defined namespace, inside a bucket. The *value* part of the *key-value pair* is the data within the object. For example, here is how an object is referenced in AWS:

```
http://s3.amazonaws.com/<bucket>/<key-name>
```

Google Cloud Platform is similar. Here is how an object is referenced in Google Cloud Platform:

```
http://storage.googleapis.com/<bucket>/
            <key-name>
```

Examinine the AWS link more closely, and notice that it includes the following:

- `http://s3.amazonaws.com`: The web address of the Amazon S3 namespace.
- *<bucket>*: Globally-unique name for a bucket. You can think of it as a root directory on a file system.
- *<key-name>*: Unique name of an object stored in a bucket. Think of this as the name of a file in a folder.

In a more detailed example, say that you have an Excel spreadsheet that is your company's ("MyCorp") quarterly earnings report. You might reference it this way:

```
C:\MyCorpData\2016_Q4_Earnings_Report.xlsx
```

If you want to save that earnings report to the cloud, here's an example URL to reference that object in the cloud:

```
http://s3.amazonaws.com/MyCorpData/2016_Q4_
            Earnings_Report.xlsx
```

MyCorpData is your bucket, and everything after the bucket is the object ("2016_Q4_Earnings_Report.xlsx").

What's in a Name?

Here's the catch. In standard file systems, you have directories and subdirectories. Suppose you store that Excel spreadsheet in the following location:

```
C:\MyCorpData\Reports\Public\2016_Q4_Earnings_
            Report.xlsx
```

Cloud object storage doesn't have the concept of a folder or directory. Everything is addressed with Bucket/Object names only; not Bucket/Folder/Sub-Folder/Object. Amazon S3 allows you to "cheat" by using "/" in the key-name of the object. For example, you can do this:

```
http://s3.amazonaws.com/MyCorpData/Reports/
          Public/2016_Q4_Earnings_Report.
          xlsx
```

In this more complicated URL, the bucket is still MyCorpData. The remainder is the key to the object — your 2016 Q4 Earnings Report. If you go to S3 using Amazon's browser-based interface, you'll see folders under the MyCorpData bucket. You can drill down inside them like in a normal desktop folder, until you get to the actual "2016_Q4_Earnings_Report.xlsx" object. However, the key-pair for that file is, and always will be, "Reports/Public/2016_Q4_Earnings_Report.xlsx." You can't refer to it only as "2016_Q4_Earnings_Report.xlsx" because "`http://s3.amazonaws.com/MyCorpData/2016_Q4_Earnings_Report.xlsx`" and the original URL refer to two different bucket/key-pair names entirely.

You must be specific when you describe an object, from the bucket all the way down to the object key-pair names. You can translate your current methods of storing and retrieving data into the cloud's object-based storage scheme. Avere Cloud NAS does this transparently as part of its data-tiering capabilities, helping manage your storage team's stress level.

Framing Cloud Compute Basics

You are probably comfortable with on-premises server virtualization. Virtualization is why the public cloud works.

The basic cloud ingredients include:

- Compute servers, running the OS of your choice
- Storage servers, which hold your files
- A network, for communication

Compute servers

Compute servers, or *instances*, are where the application magic happens. Physical servers in the cloud are grouped into massive multi-server pools of CPU, memory, and disk. Instances are created by slicing off X number of CPUs, Y number of RAM, and Z number of virtual disks from the pool.

The piece of software that allows server resources to be parceled out is called a *hypervisor*. Think of a big loaf of bread being the pool of all the servers' CPUs, memory, and disks. A slice of bread from that loaf represents a virtual server instance, which behaves like a single, physical server.

When you request a server to be created in the cloud, you can describe what operating system you want it to run, how much memory you need, and the number of CPUs. You also can choose from varying types of standard setups, depending on what kind of applications you're planning to run.

Storage

Although it can vary, *cloud storage* is always a type of shared storage that can be connected to multiple servers. This ability to share makes it easy to slice off what you need and give it to whichever server needs it. When you're done, you disconnect it and give it back to the pool for other servers to use.

Storage in the cloud can have different characteristics, such as redundancy and a guaranteed I/O rate. You'll pay more per gigabyte for highly available and/or faster storage.

Network

Cloud networks are *virtual.* You don't have to run wires to get servers and storage to communicate. The servers are connected to the network when they're built, and everything else, such as IP addresses, routers, DNS entries, and firewall rules are all handled either automatically or customized with management software by the administrator.

Looking at Cloud Types

AWS, Google Cloud Platform, and Microsoft Azure are the three public cloud gorillas. You can mimic public cloud behavior with on-premises virtualization. You can also create a *hybrid cloud* by combining local and cloud resources.

- **Private clouds:** These are virtualized environments you build yourself. You are responsible for all the equipment, which necessarily means you have physical scaling constraints.

- **Public clouds:** This book focuses on the IaaS aspect of the public cloud. Offered by AWS, Google Cloud Platform, Microsoft Azure, and others, public IaaS provides pay-per-use storage and compute at massive scale.
- **Hybrid clouds:** Enterprise companies typically have a significant investment in existing infrastructure. By combining local resources with those available in the public cloud, you can maximize your existing infrastructure investment and incorporate the cost savings and flexibility of public cloud.

Moving data from a company's local environment to the public cloud isn't trivial, but using Avere Hybrid Cloud NAS to bridge the gap between them is an efficient way to do so.

Transitioning into the Cloud

You're thinking, "How do I get started, while minimizing impact?" Don't worry. The Avere Hybrid Cloud NAS solution helps you optimize your current infrastructure investment while gaining the benefits of cloud.

Figure 2-1 gives you a sneak peek into the future, where you can bridge the divide between on-premises and cloud infrastructure to create Hybrid Cloud NAS.

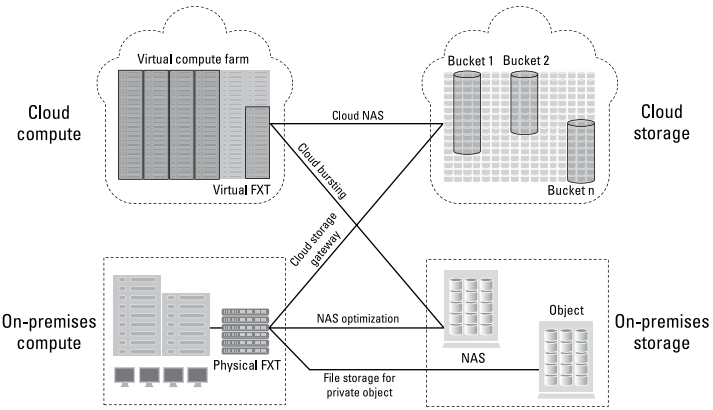


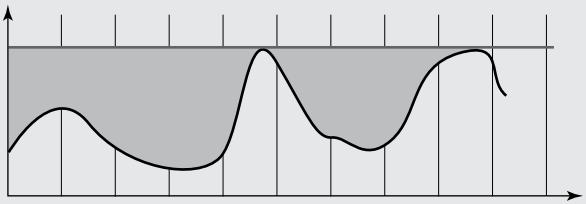
Figure 2-1: Imagine a world where storage and compute can interact, regardless of location.

Leveraging Excess Capacity

One neat thing about the cloud is the sheer size of it. The footprint of a typical IT organization is almost a rounding error in a cloud provider's capacity plan. Excess cloud capacity leads to an interesting capability — the temporary server market. AWS calls servers of this type "spot instances," while Google Cloud Platform refers to them as "preemptible virtual machines."

Cloud providers don't want customers to experience performance issues. That's why they maintain more capacity than their customers use on a daily basis. Big batch jobs combine with spiky workloads on the same infrastructure. The capacity sucked up by a big batch job is idle outside the batch window.

Consider the diagram in the accompanying figure. The horizontal line illustrates total available capacity, the fluctuating line illustrates the variable customer workload, and the shaded area illustrates idle capacity.



Variable workload with wasted capacity

To make that capacity productive, providers make it available for use — with a catch. Whenever a customer who has reserved capacity needs to run a workload, the server instances currently occupying that capacity get turned off. Although the thought of someone turning off your server can be initially terrifying, it's great for processing stateless workloads.

You're probably thinking, "With a catch that big, there must be a benefit!" Spot/preemptible instances are more than 70 percent off the list price of a regular instance! That's a big deal, if you have applications that can handle having their host servers turned off at a moment's notice.

Chapter 3

Leveraging Cloud Storage

In This Chapter

- Linking your on-premises storage with cloud storage
- Unlocking the power of the cloud
- Connecting cloud compute with local storage
- Activating the hybrid cloud

In Chapters 1 and 2, I discuss network attached storage (NAS), cloud compute, and cloud storage. This chapter explores Avere's innovative Edge filer technology. These advanced pieces of engineering can be used to take advantage of cost-effective cloud storage while maximizing your investment in on-premises infrastructure. Combining local and cloud infrastructure into a hybrid gives you the best of both!

Protecting your Investments

Historically, enterprise companies delivered IT services from physical data centers and co-location facilities. IT departments are often thought of as cost centers. They face ongoing budget pressures while dealing with increasing service demands. If you're used to thinking of data centers in terms of megawatts, the public cloud represents a flexible, cost-effective shift.

Many companies have an architecture with local compute servers talking to a local NAS. Concerned with latency, IT directors and storage architects look longingly at the cost-effective yet unfamiliar object storage provided by services like AWS, Google Cloud Platform, and Microsoft Azure. The Avere FXT Edge filer series helps you bridge the gap.

As shown in Figure 3-1, an FXT Edge filer is a physical appliance that is set up between your on-premises compute and storage environments. Filled with intelligent software that can cache and tier data, loaded with gigabytes of RAM and solid-state drives (SSDs), Edge filers provide your compute environment with the fast access to storage your applications need and your business demands. Smart caching and pre-fetch algorithms ensure that active data is kept as close as possible to compute to help optimize application performance. Capable of being deployed in clusters from 3 to 50 nodes, FXTs can be horizontally scaled to provide millions of IOPS.

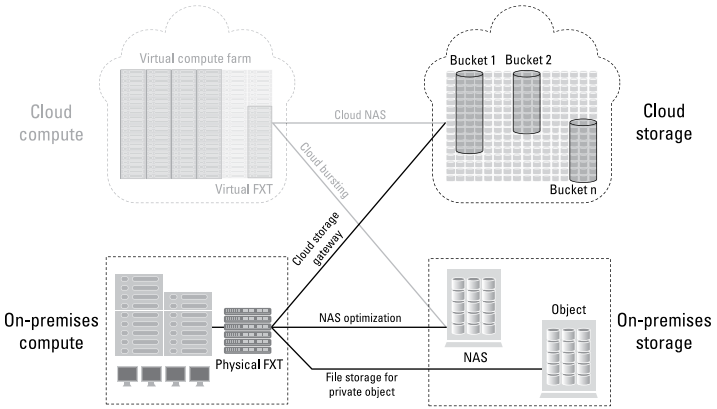


Figure 3-1: Avere FXT Edge filers provide a global namespace that spans local and cloud storage.

An FXT cluster can also attach to the public cloud and provide on-premises compute with a global namespace that spans both local and cloud-based storage. Using application programming interfaces (APIs) to connect to local or cloud object storage, the FXT can present both Network File System (NFS) and Server Message Block (SMB) to your compute environment. With its algorithms and fast media doing the heavy lifting, the FXT appliance addresses the latency problem typically associated with the public cloud.

Interested in instant geographic diversity and disaster recovery for your local NAS? I thought so. Baked into the FXT is a feature called FlashMirror, which can mirror data between local and cloud storage environments. Trying to migrate from a local NAS to the cloud? Storage engineers can use FlashMove, a data mobility tool built into the FXT, to non-disruptively migrate data between local and cloud storage.

NAS with Cloud Scale

Highly performing teams are made up of people with differing perspectives and points of view. Similarly, introducing physical FXT Edge filers as components of your overall storage architecture lets you maximize the best features of local NAS and cloud storage. Combining cloud storage with NAS lets you solve storage challenges that can't be fixed by local NAS or cloud alone. Speed, latency, cost, and complexity are the main drivers in managing storage. By combining the features that make NAS or cloud a good solution on their own, you can creatively deliver business value in entirely new ways.

Hitting the gas

Speed! We all crave it. Humans have been obsessed with speed since the dawn of time. If it can move, people will end up seeing how fast it can go: sailing ships, bicycles, motorcycles, cars, planes. This section explores how you can optimize storage performance for your local compute environment using Avere's Hybrid Cloud Edge-Core architecture.

Fronting your existing NAS storage with an FXT filer ensures speedy storage performance for your compute farm. Separating performance from capacity, Avere's Hybrid Cloud architecture delivers both more efficiently. Edge filers supply the performance your applications need, while abstracting the slower performance of cloud storage, as illustrated in Figure 3-2.

Edge filers provide high-performance and low-latency data access to your compute farm. This is possible because of three distinctive features of the FXT:

- Dynamic caching, pre-fetching, and tiering algorithms run inside Edge filers to automatically move the hottest data blocks from the slower disks of the Core filer(s) to the Edge filer's fast media.
- As a full-featured file server, Edge filers locally terminate all metadata, read, and write operations, avoiding the latency to the Core filer(s), which can be very high, especially if they are connected over a WAN.
- Clusters of Edge filers work together as a group, distributing file requests from clients and dynamically scaling to handle the I/O workload.

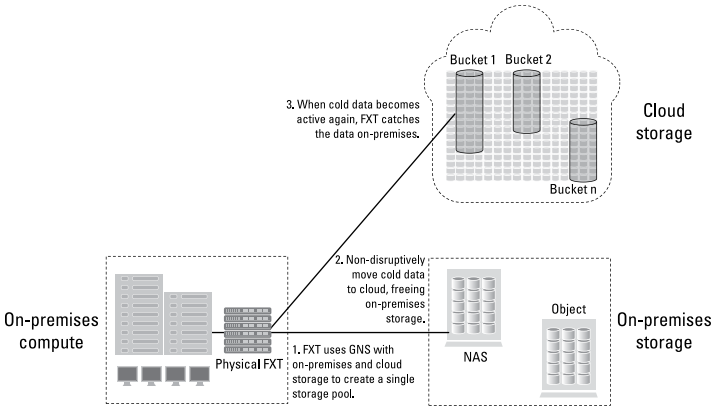


Figure 3-2: Avere Hybrid Cloud Edge-Core architecture presenting public object, private object, and legacy NAS in a global namespace.

With an Edge cluster on the front lines providing fast access to hot data, the demands on the core on- and off-premises storage subsystems are reduced. The hybrid approach gives you better optimization of performance, capacity, and cost.

Saving at the pump

The explosive growth of unstructured and structured data is placing immense pressure on storage costs. The public cloud offers the promise of financial relief in the form of cheap, pay-as-you-go storage. Avere Hybrid Cloud architecture lets you realize that potential savings.

Although data has value, its storage comes at a cost. If you look at the number of times a piece of data is accessed, you can get a good sense of how valuable it is to your business. By using algorithms to calculate the usefulness of data based on activity levels, Edge filers can accurately figure out a specific piece of data's value. Edge filers help you save money by automatically moving low-value, infrequently accessed data to the cheapest storage location — on-premises or in the cloud. Because this happens automatically, the Edge filers are constantly working to make sure the most valuable, frequently-accessed data is closest to the systems that need it.

Once comfortable with the concept of using Edge filers to buffer the I/O demands of mission-critical applications, you can turn your attention to cutting the cost dragon down to

size. What's the cheapest place for bulk storage for your company? Local NAS? AWS? Google Cloud Platform? Microsoft Azure? By sizing your Edge cluster appropriately, you can take advantage of the operationally cheapest long-term storage, while knowing your applications will meet their performance targets. If one public cloud becomes significantly cheaper than another, you can rest easy knowing that switching long-term storage vendors won't change the way your line-of-business applications operate.

Adapting existing applications to use object storage is hard. Having vendors change their applications is even harder. Edge filers abstract that pain, letting your applications use familiar NFS and SMB interfaces. This means your application teams can spend more time on features that give your business an edge, and less time worrying about storage interface protocols. Without sweeping application architecture changes, Edge filers give you flexibility and help your bottom line.

Using the public cloud and concerned about IOPS? Consider putting a virtual FXT (vFXT) cluster, a software-only version of the FXT that runs in the compute clouds offered by AWS, Google Cloud Platform, and Microsoft Azure, in front of S3 storage offered by the same providers. No need to sweat cost and maintenance associated with managing lots of expensive block storage in the compute cloud — simply front cost-effective S3-based storage with a vFXT cluster, cache only the hot data in the compute cloud, and scale it up to your needed performance level!

If you come to the common conclusion that it is more cost-effective for your business to let AWS, Google Cloud Platform, or Microsoft Azure operate your physical storage infrastructure, Avere's Edge filers give you the performance your applications demand while letting you choose the cheapest storage backend.

Making things simpler

FXT Edge filers offer the Avere Global Namespace (GNS), simplifying data management and access in a geographically diverse storage environment. Avere GNS spans public object, private object, and legacy NAS, all from heterogeneous vendors (see Figure 3-2).

Need to access data from Google Cloud Storage with your local RedHat boxes? This challenge faces many organizations during cloud transition. Thankfully, your FXT cluster handles the object storage to NFS protocol translation. With Avere FXT-powered GNS, compute clients have a single mount point on the FXT cluster. Freedom from storage protocol worries gives your applications access to all the data, regardless of where it is ultimately stored.

Avere GNS implementation doesn't remap your existing file systems. No need to worry about vendor lock-in, because the GNS is simply a reflection of the namespace you create.

Easy Data Migration

Avere FlashMove is a piece of data management software that runs in every FXT Edge filer. This software makes optimizing heterogeneous storage environments a breeze.

Thinking of switching local NAS vendors? No problem! Because Avere supports multiple legacy NAS vendors, including EMC and NetApp, migrating from one to the other is as simple as executing a FlashMove command. Want to trade power, cooling, screwdrivers, and racks for the public cloud? Edge filers can FlashMove your data for you without changing the storage interface for your applications. No application rework required. Easy!

As surely as the sun rises in the east, you can be sure that your storage team will have to perform data migrations. The normal hardware lifecycle and switching NAS vendors are just a couple of examples. FlashMove eases the pain on your storage team by reducing the effort required, transparently and non-disruptively moving data from your original storage subsystem to your new storage target. Users and applications continue to see live data while FlashMove handles data migration in the background.

With FlashMove, you don't need to halt applications or suspend access to data during migrations. Edge filers continue serving active data to people and application servers, while the FlashMove software transparently migrates data from source to target storage subsystem. You can load-balance live data across existing systems, transparently archive to slower

and cheaper storage, decommission old storage, add new storage vendors, and take advantage of public cloud storage, all with a few simple commands.

You can perform a FlashMove on live data without disrupting the operations of your existing applications! Non-disruptive NAS-to-cloud, cloud-to-NAS, or cloud-to-cloud data migrations are at your fingertips.

Protecting Your Data

Running a data center is an expensive proposition. Losing the data on which your company relies is an unacceptable proposition. Nobody wants to tell the CEO about data loss. Making copies within a single geographic area won't help when a natural disaster rolls in — just ask any company that had its operations interrupted by Hurricane Sandy in 2012. Having physical access to your data center doesn't matter if that data center is literally under water.

Building a data center in a separate part of the country is cost-prohibitive for most companies. Co-location facilities are only slightly cheaper. Frequently, the public cloud holds the most cost-effective solution to this challenge, offering cheap object storage in multiple geographic regions across the globe. While you yearn to take advantage of this capability, Avere helps you get there by building FlashMirror into every Edge filer.

Suppose you're happy with the cost and performance of your local NAS, but your current configuration doesn't meet your organization's recovery time objectives (RTOs) or recovery point objectives (RPOs). With FlashMirror, your FXT Edge filer comprehensively protects your critical data by mirroring it in AWS, Google Cloud Platform, or Microsoft Azure. Instead of incurring the cost of establishing multiple physical remote data centers, you can rely on the public cloud. After all, public cloud suppliers employ seismologists when selecting sites for their respective data centers!

When considering cloud object storage, you can't compare apples and oranges. Every time you write to cloud storage provided by AWS, Google Cloud Platform, or Microsoft Azure, the data you write is replicated across multiple data centers. Can you say that about your local NAS?

From the moment you execute the FlashMirror command against a source directory, your FXT cluster creates a baseline copy in secondary storage of the source directory. Like a legacy NAS mirror, FlashMirror ensures that any change made in the primary storage is reflected in the secondary location.

Prior to the next scheduled disaster recovery (DR) review at your company, talk with your storage architects about how FlashMirror can help achieve the RTOs/RPOs management is asking for. You can improve your data protection and provide for business continuity without exorbitant additional expense.

Connecting to the Cloud

The public cloud offers practically infinite, pay-as-you-go compute and storage capacity. With such massive economies of scale and the possibility of limiting waste, traditional companies are rapidly developing plans to take advantage of the capabilities the cloud offers. The challenge is, most cloud storage is object storage — how do your existing applications connect to it? What you need is some kind of doorway to the cloud. A cloud gateway!

Avere FlashCloud is a feature baked into every Edge filer that provides cloud gateway capabilities. With FlashCloud, integrating object storage with legacy NAS into a GNS is a snap, providing scalable performance via familiar NAS protocols. This abstracts your staff from the unfamiliar application programming interfaces (APIs) used with cloud object storage.

GNS lets enterprises adopt object storage at the pace of their choosing, while offering the flexibility to store data wherever it makes the most sense. FlashCloud is what Avere calls the piece of technology that abstracts object storage interaction by providing enterprise-class NAS functionality protocols, including NFS and SMB. Working with cloud and local storage, Avere's filers give you the optimized performance, capacity scaling, and reduced total cost of ownership you are looking for.

Changing the Center of Gravity

With the attractive pricing of cloud storage, it is quickly becoming the location of choice for massive data sets. The center of gravity is swinging away from local NAS toward the titans of AWS, Google Cloud Platform, and Microsoft Azure.

In the life sciences field, genomics research creates lots of data. How much? One company has in excess of 10 petabytes — more than 10,000 terabytes! Imagine building and maintaining a NAS to hold that much data. Now ponder continuing to build a NAS to store the data generated by the increasingly-affordable genetic sequencing instruments.

Instead of storing all that data locally, one research institute has been using the Avere Edge filer as a cloud gateway to move multiple petabytes of data out to the cloud. With this approach, the institute gets maximum value from its existing NAS while embracing cloud storage. And, when the cloud-stored data gets active again, the Avere Edge filer caches the hot data on-premises to give researchers low-latency access as if all the data were still stored on premises. This allows the institute to focus funds on research instead of on an immense local NAS infrastructure.

Meanwhile, over in financial services, companies are dealing with massive amounts of data. Hedge funds are on a never-ending quest to improve algorithms to maximize financial performance. In order to validate whether or not those tweaks will work, companies backtest the updated algorithms against historical data. Storing decades of detailed stock market transactions on a local NAS isn't worth the cost. The data doesn't change.

Using a cloud gateway allows static data sets like these to be stored out in the public cloud. When the financial math whizzes come up with a new algorithm to test, they don't have to do anything special. The historical stock market data comes transparently down from the cloud via the cloud gateway.

Launching into Space

Storage exists in the clouds — but what about beyond? What's out in space? The knowledgeable folks at NASA's Ames Research Center can tell you all about it. Because they have stewardship over such an enormous trove of imagery data, storing it locally was becoming a major cost center.

Leveraging FlashCloud and FlashMove, the Ames Research Center continued using NFS and SMB, embraced cloud object storage, and slashed overall costs without compromising operational performance.

Humans don't have data centers in outer space — yet! By reallocating funds from IT into research, maybe we'll get there someday!

Making Things Better

Faster, cheaper, more robust, scalable, and easier to operate — that's my definition of making IT operations better. Let's explore robustness and scalability for a moment. Traditional design practice to address these issues is with a highly available cluster. In keeping with industry-accepted best practice, the vFXT and FXT Edge filers can be clustered in a high availability configuration. This helps ensure that access to storage is resilient, robust, and scalable.

Avere Hybrid Cloud provides the flexibility you need to maximize existing capital investments while taking advantage of variable compute and storage in the cloud. This lets you focus precious capital dollars on improving the robustness and reliability of architectural components that aren't candidates for moving to the cloud — network links and core network infrastructure, for example. This flexibility lets you adopt the cloud at a pace that is right for your organization.

There now, isn't that better?

Chapter 4

Bursting Compute to the Cloud

In This Chapter

- Exploring benefits
- Noodling over usage models
- Zapping constraints
- Opening new doors to operational excellence

This chapter explores many of the non-technical ways in which companies can benefit from cloud bursting. *Cloud bursting* augments your local compute environment with capacity from public cloud vendors. The ability to cloud burst truly changes the paradigm of how to think about compute and storage.

Benefits

Technology is flashy and fun, but it doesn't mean a thing if the cash register doesn't ring. Ultimately, IT is about enabling the business to create and deliver value. This section explores some of the non-technical benefits of cloud bursting.

Chapter 3 talks about how existing enterprises have big investments in local network attached storage (NAS). Similarly, companies have major investments in compute. The cost is typically an aggregate of physical servers and virtualization software license fees. Managing the lifecycle of physical assets is slow, compared with the ability to spin up a virtual server in minutes in the public cloud.

Some companies are quite satisfied with their existing storage infrastructure but find themselves constrained by compute cycles. Cloud bursting lets these companies keep their data locally while adding compute power at peak times, as illustrated in Figure 4-1.

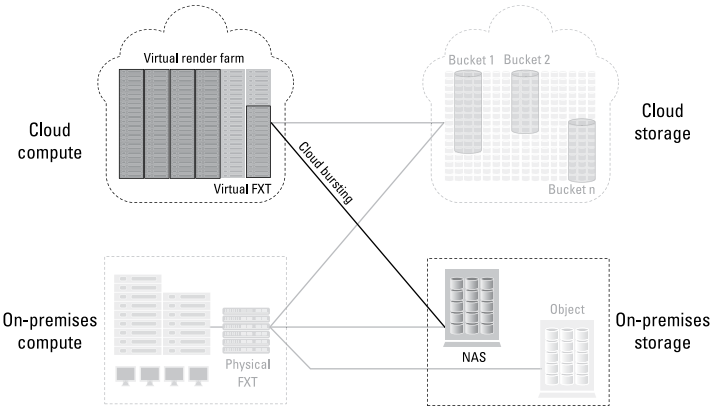


Figure 4-1: Using cloud computing to access on-premises storage.

All about your business

In most companies, IT is not the main reason why the company is in business. IT exists to serve the strategic objectives of the business. Sometimes, IT pros can get so caught up in the joys of technology that they lose perspective of the fact that the entire point of IT is to serve the mission of the company.

As compute and object storage becomes increasingly commoditized, AWS, Google Cloud Platform, and Microsoft Azure can realize massive economies of scale. If you think your company is in a position to design a more efficient, more cost effective data center, you should seriously consider getting into the public cloud game.

All joking aside, the point is that the industry will continue to see the global consolidation of data centers. Let the pros at AWS, Google Cloud Platform, or Microsoft Azure handle the racking and stacking. Let them worry about power and cooling. All the manpower and money being used to maintain large local compute farms can be redistributed. When demand spikes, you can rest easy in your ability to simply add compute capacity, courtesy of cloud bursting.

Letting go of the hardware allows you to focus on the reason you are in business. Go ahead; unleash your artistic creativity for that next movie. Focus on the science required to achieve that medical breakthrough. Refine and rework your analyses to gain new insights. And rest well at night, knowing that you can instantly double your compute capacity, if you need to, by firing up a fleet of servers in the cloud without reducing storage performance.

Instead of rushing through your next-generation architecture, you can create a plan to meet the timelines your business demands.

Becoming more agile

Without the compute constraints of your local infrastructure, cloud bursting lets you think differently about how to create solutions. The printing press revolutionized the distribution of ideas — cloud bursting is revolutionizing how quickly you can transform ideas into reality!

While developing predictive models based on stock market data, a leading financial services firm came up with an idea that could have a significantly positive impact on its portfolio performance. Thinking the idea had merit, the firm needed to prove its theory using real data. Getting the data wasn't a problem — it was all sitting on local storage. Figuring out how to process it quickly and cost-effectively was the challenge at hand.

The firm's quantitative analysts use the Monte Carlo method to prove out its financial models. Understanding that Monte Carlo simulations rely on repeated random sampling to get results and are incredibly compute-intensive, the firm had to figure out how to make that happen. The burning question was: How could the company access the compute power to see if the analysts' theories were worth changing the investment approach?

Enter cloud bursting. The firm fired up 30,000 cores in the Google Compute Engine to perform its risk analysis. To minimize latency and let those cores access the required data, the firm used an eight-node virtual FXT (vFXT) Edge filer from Avere to provide the required storage performance.

Think about that for a second — 30,000 cores! Imagine what that would have required in the firm's own data center. They would have placed an order with a server manufacturer. The servers would have been built and shipped to a data center. Because 30,000 servers generate lots of heat, cooling would have been upgraded. At the data center, the servers would have been uncrated, racked, plugged into power, and hooked up to the network. That's assuming the data center had enough space — if not, a pricey data center expansion would have occurred. After all that work, the application would have been installed. Finally, simulations could have begun. How long would that take? A month? Longer?

With cloud bursting, that lengthy process doesn't have to happen. By using the public cloud and providing access to data with an Avere vFXT Edge filer deployed in the cloud, the firm could run simulations quickly and cost-effectively, paying only for the compute it used. Even better, the simulation models could be updated directly on local, on-premises storage by the firm's quantitative analysts because of the cluster's global namespace capability. The result was a significant improvement in the performance of the firm's investments.

What does the future hold? For this particular firm, more intense simulations, run more frequently. How intense? The firm has plans to move from 30,000 to 100,000 cores. *100,000!* That's a pretty serious number. The firm is confident that the cloud will be ready with the compute power needed, and that the storage cluster can provide the performance required by scaling horizontally.

Commodity usage

Is compute capacity really different from electrical power? Unless you have unique needs for specific hardware components, the answer is no. You simply have code that needs to execute somewhere — you don't care where that is.

One path to maintaining your competitive edge is by shortening cycle times. The faster your company can react to changes in customer demand, market forces, or regulatory constraints, the better positioned you are for sustained success. Sometimes, thousands of hours of processing time can stand between you and the next release of your product. This

is true in many industries, whether producing the next smash hit movie, designing the next cancer-fighting drug, crunching research data, or developing the next championship-winning race car. What if your project is running behind schedule? This is where cloud bursting can give you an edge.

You know that the cloud has almost limitless compute capacity. With a few mouse clicks or a bit of scripting, you can spin up thousands or tens of thousands of servers in AWS, Google Cloud Platform, or Microsoft Azure. Bang — you can quickly change the compute power you have at your disposal, as easily as turning up the volume on your smartphone. With the ability to scale storage access to keep pace with the cloud's augmented compute power, your project will be back on track.

You know what you can't forget to do after you've ground through that mountain of data? Shut those cloud servers down, just like shutting off the lights when you leave your office. Remember, in the public cloud, you pay only for what you use. The cloud lets you dial in exactly the right amount of compute performance, exactly when you need it. Click. Goodnight.

Being early is better than being late. Incorporate the ability to cloud burst into your product planning cycle. Use the knowledge that you can bring tens of thousands of machines online in a crunch.

The end of refresh cycles

Think of how much effort you spend just "keeping the lights on" in your compute farm. You fret over firmware, agonize over air conditioning, and panic about power consumption. That's only the beginning! You ponder physical plant expansions and get ulcers about geographic diversity. When is that faster CPU coming out? What percent of my compute infrastructure do I refresh on an annual basis?

Imagine never having to do that again.

Think about what you could do if you didn't have to budget, plan, and execute hardware refreshes. Imagine getting the latest CPUs just by rebooting a server. With cloud bursting,

you can replace your nightmares with dreams of what you can do with all your newfound time.

Burst Models

Companies are approaching cloud computing in several ways. This section explores them.

Temporary Usage at Peak Times

Consider the following questions: Is your business affected by seasonality? Do you have predictable swings in demand? What about spiky, unpredictable demand? If you answered "Yes" to any of these questions, your organization is a prime candidate for cloud bursting's model use case. As I've mentioned, you pay for cloud compute or storage only when you use it. Burst out to the cloud as the variability happens to sustain the performance your customers expect.

What about non-production systems? Here's some simplified math. Say you have an equal number of development, test, and production systems. Do your development and test systems need to be on as often as your production systems? Do these systems need to run at night, when your staff is peacefully slumbering? Odds are, probably not. Embrace cloud bursting by moving these temporary environments into the cloud and shutting them down when they are not being used.

All-in!

Are you beyond the point of temporary usage? Are you ready to turn in your wrenches and screwdrivers and commit to a cloud-centric compute environment? Do you want to migrate to the cloud but run into snags, hampered by proprietary software or legacy applications that don't understand object storage? The hybrid cloud facilitates this transition by allowing you to move your data where and when you are ready, with the storage protocols you already use, at a pace you control. Think about it — you don't need to make changes to your application stack to start realizing the numerous benefits of cloud computing and storage.

Zero Gravity

Moonbot Studios is a media and entertainment company based in Louisiana. Far from the glow of the Hollywood sign, the buzz of Silicon Valley, and the population density of New York, the talented artists at Moonbot are passionately striving to produce Academy Award-winning work. And succeeding.

In an industry filled with powerhouses like Pixar and Dreamworks, capital expenditures are of paramount importance. Rendering a modern movie requires an incredible amount of computing power. Individual frames can often take hours to render, with some films consuming over one *petabyte* of storage. That's a lot of data! Leading visual effects companies have built massive compute farms in excess of 80,000 cores across thousands of nodes. Moonbot enters the fray with a grand total of 36 rendering nodes.

With only 50 people, Moonbot doesn't have the luxury of building and operating a massive physical render farm. Cloud bursting to the rescue! With some clever scripting on Google Cloud Platform's Compute Engine, Moonbot can fire up more than 150 additional compute nodes in about ten seconds.

Meanwhile, most of its data is stored locally. With more than 100 terabytes of data, copying from its local NAS to the cloud would not have been acceptable from a performance standpoint. By implementing a virtual Avere vFXT Edge filer, those Google Compute Engine nodes can be productive as soon as they come up.

Moonbot can focus on what it does best — recruiting great artists and producing great content. The ability to cloud burst makes IT stuff easy.

Data Can Be Anywhere

For traditional companies, most data lives on-premises today. The trend is for that data to migrate to cloud storage in the years to come. Commonly, applications are hamstrung by data location. Migration strategies can hinge on data, with applications unable to move until, for example, a database moves.

With a hybrid cloud design, your cloud-based applications can access data from anywhere, at NAS speed, even if it still resides on premises. While taking advantage of this unique design pattern, you can start planning when to migrate your data to complete your transition to the cloud. Meanwhile, you

can offload your cold on-premises data and start using cloud storage today.

Speed of Flash, Cost of Disk

There's a saying in racing: "Speed costs money — how fast do you want to go?" For your applications to perform fast enough to meet the demands of your business, you need to ensure that your compute has consistently fast access to storage. It is possible to get this level of performance when you use a physical appliance to front cloud storage, or a virtual appliance to front on-premises storage to your cloud compute. With an appliance handling active requests, the bulk of your data can be offloaded by the appliances onto whatever storage is most cost-effective for your organization.

Okay, so cloud object storage is cheap, but what about the speed part of the racing question? That's where you have to rely on the software built into physical and virtual Avere FXT Edge filers. With intelligent algorithms, appliances can identify hot files before you even know they exist. Applied appropriately, these algorithms automatically pull hot files into the filer's RAM. As the file cools off, it is automatically tiered to progressively slower media, passing through the filer's SSDs before becoming a tiny ice cube in the corner of the cheapest object storage you can find.

If a file ever looks like it's going to get hot again, pre-fetch software comes into play, pulling the file back up the storage tiers, so your applications never know the difference or experience a dip in performance.

If your car didn't have gauges, you wouldn't know how fast you were going or how much gas you had. Similarly, when evaluating solutions for cloud bursting, don't underestimate the importance of a solid dashboard to give you a perspective into how your system is performing. Make sure that you can see hot clients, files, current and historic IOPS, throughput, and latency at a glance. This helps give you the confidence to know that your applications are benefiting from the speed of flash, while you pay for the cost of disk.

Chapter 5

Embracing Cloud NAS

In This Chapter

- Exploring possibilities
- Navigating worries
- Zero constraints
- Optimism reigns

In this chapter, I explore how you can use Cloud NAS to easily scale both compute and storage. Forget worrying about messy protocols! Focusing on giving your business what it needs, when it needs it, is what Cloud NAS is all about.

As Elastic as your Imagination

Your team has spent decades building expertise in providing your company with the technology it needs. You've built relationships, gone through forecasting exercises, and managed capacity plans. In a world of horses, you've created the best possible horseshoes. And yet, the rate of competitive change facing your business is increasing.

You've dabbled in the cloud. Maybe you got started with storage. With your network attached storage (NAS) overflowing, you found relief in the vast ocean of cheap, available cloud storage. Perhaps you've explored cloud bursting, using the readily-available compute capacity from AWS, Google Cloud Platform, or Microsoft Azure.

Reflect on how truly challenging it is to operate a world-class infrastructure. Many of the leading server and storage vendors have tried to compete with the three cloud titans, and failed. Do you have the resources to compete with them? What if you just — didn't? What if you opted out of on-premises

equipment? What if, instead of taking on the really hard work of building a highly redundant, available infrastructure, you simply chose to stand on the shoulders of giants? AWS builds its own power substations. Google manufacturers its own servers. Why even try to compete? Put the billions of dollars these titans have invested in infrastructure to work for you! Go cloud native!

The top of Figure 5-1 is what I'm talking about. Migrate your data to the cloud, and shut down your local storage. Spin up virtual machines in the cloud and jettison your local compute environment. Better yet, if you don't have a legacy on-premises infrastructure already, don't build it at all!

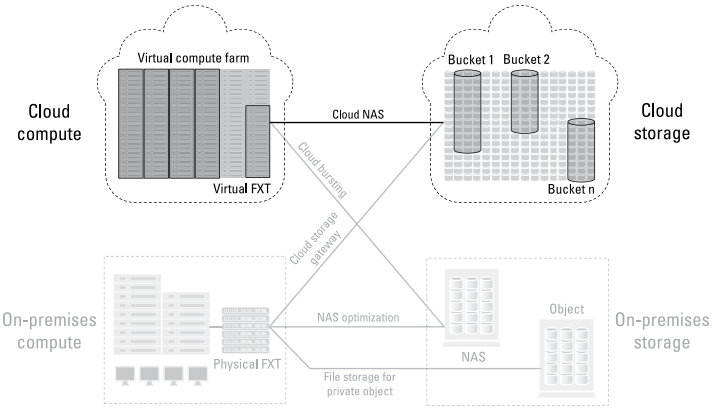


Figure 5-1: Do away with your on-premises compute and storage and fully embrace the cloud.

Imagine combining all your knowledge and expertise with the agility of a cloud-native design philosophy. Imagine being able to pivot your infrastructure as fast as your business model. Imagine that instead of being looked on disdainfully as a cost center, IT is viewed by your executive leadership as the truly strategic advantage that it is. By truly embracing the offerings from AWS, Google Cloud Platform, and Microsoft Azure, that is exactly what you can do.

Jumping in with both feet

Looking at your existing applications, you know where the performance bottlenecks are. Maybe you're consistently

CPU bound. Maybe it's storage, or memory, or network. Ask yourself — is the bottleneck consistently one thing?

Is your load consistent? Perhaps it's utterly predictable. More likely, you manage batch workloads, spiky seasonal demands, consistent growth, and maybe even spikes brought on by attention from the media. With the ability to alter your compute or storage capacity within seconds, you'll soon appreciate why even mega corporations like General Electric have made a splash by diving into the cloud.

Are you worrying about the effort it would take to retool your line of business applications to be cloud aware? Take comfort in the fact that Cloud NAS solutions exist to help your applications talk to storage using their current protocols.

Freedom from physical infrastructure

I know what you're thinking. Put all my compute *and* storage in the cloud? Madness! Talk about vendor lock-in! Relax just a moment and think. Don't worry about vendor lock-in. Focus on automation. Pick a tool — I don't care which one. Chef, Puppet, Salt, potayto, potahto. Pick a tool and focus on turning infrastructure into code. You'll come out ahead.

Still worried about lock-in? Pick a tool to abstract the nuances of each cloud provider. You can get close to being able to script the entire creation of your datacenter — in any of the big three cloud providers. Or even in more than one of the big three!

Imagine the freedom to design a technical solution that is unshackled by the constraints of physical infrastructure. Geographic diversity and disaster recovery become logical exercises you can implement in hours.

Freed from physical infrastructure, it takes seconds to spin up some servers or augment your storage. With no hardware to touch, you can get up and running with just a few clicks. Just remember the knowledge Captain Kirk shared with Spock in *Star Trek II: The Wrath of Khan*, "Hours instead of days! Now we have minutes instead of hours!" Being free from physical constraints makes that possible.

Cloud NAS Case Studies

Companies are quickly realizing the many advantages of getting out of the data center business. This section explores a few examples to get your creative juices flowing.

Genomic analysis

H3 Biomedicine (H3) is a Cambridge, Massachusetts-based clinical stage biopharmaceutical company with a bold vision. H3 is combining cancer genomics, synthetic organic chemistry, and tumor biology to create cancer treatments that are unique to genetically-defined patient groups. With a mission to individualize the cancer fight, the genomics analysis processing requirements are immense.

Founded in 2011, H3 is very much a startup. It has no tolerance for allocating assets that don't directly contribute to advancing H3's core mission. Deeply centered on scientific research, H3's computational biologists need to use compute power the same way its cell biologists use microscopes. Facing 30 percent year-over-year data growth, keeping pace using local infrastructure is impractical. To solve scale challenges while giving bioinformaticians access to the tools they need, H3 turned to the public cloud for horsepower, using a hybrid cloud approach to ensure storage performance.

While the scientists' puzzle is how to target cancer, H3's research computing architect needs to make more than 80 terabytes of data accessible to a multitude of compute nodes. Instead of building a massive compute farm, H3 uses AWS Elastic Compute Cloud (EC2). Just like rendering for movies, the nature of H3's workload requires lots of processing power for intense periods. This is a perfect match for the cloud because many of the jobs scientists run are easily parallelized.

One advantage H3 gets from its design is a greatly simplified architecture and associated operational procedures. Instead of dealing with EC2 block storage, H3's EC2 and local compute instances simply interact with the Global Namespace presented by its storage cluster. Instead of designing a data synchronization engine to shuttle data back and forth between

local storage and AWS, H3 can focus its assets on what differentiates its business the most: scientific research.

H3's on-premises compute and storage will go through a natural lifecycle. When these components reach end-of-life, the intention is to migrate more and more compute and storage into AWS. With an Avere vFXT Edge filer, H3's Linux tools and applications can talk NFS instead of being rewritten to use object storage. Financially, using a virtual filer cluster to handle the applications' performance lets S3 handle the bulk storage without managing large amounts of AWS Elastic Block Storage (EBS), saving even more money.

With the ability to get compute and storage at will, H3 has all the technology weapons it needs to fight cancer.

When using cloud storage providers, putting data into the cloud is free. Getting data out of the cloud is another matter entirely — you pay retrieval and network fees. When putting a frequently-accessed dataset in cloud storage, it's highly recommended you deploy the application as well in the provider's cloud compute.

Legal analysis

Has your company ever been embroiled in a legal battle? Have you ever had to deal with legal holds and e-discovery? If you haven't, consider yourself lucky. If you have, you know that collecting the information for just one e-discovery request can occupy hundreds of gigabytes of space. Apart from generating gobs of infrequently-accessed data, the processing power needed to aggregate that data is intense. Best of all, the data that is generated needs to sit around for a long time.

Many e-discovery tools run on Microsoft Windows. Presenting petabytes of cheap object storage as SMB makes this kind of workload a textbook case for Cloud NAS. Because pushing data into the cloud doesn't incur any additional charges, you can cost-effectively create a solution that addresses the business need of collecting and storing e-discovery data.

Couple that storage with the ability to dynamically bring compute capacity to bear based on e-discovery queries. Now you've created a cloud-native solution for handling the

messy business of e-discovery requests. That is exactly what a leading consulting company did, combining the power and flexibility of cloud compute and storage platforms to give the company a competitive edge for winning contracts.

Super-secret workloads

Every company has its secrets — things that provide a strategic competitive advantage, from the recipe for Coke to aerospace engineering details about aircraft. Do you think the United States intelligence community has secrets? Most certainly it does. Built on public cloud technology, agencies like the CIA have access to cloud resources that are completely isolated from the Internet.

Perhaps your company works on U.S. government contracts. With vendors like AWS offering an environment that complies with International Traffic in Arms Regulations (ITAR) and Federal Risk and Authorization Management Program (FedRAMP) requirements, you can design a Cloud NAS solution to do the research you need, all while being compliant with sensitive data handling standards.

The CIO of the CIA has spoken publicly about the need to outsource technical innovation. If the CIA is comfortable getting out of the data center business, odds are that you have nothing to worry about.

The Undiscovered Country

In *Hamlet*, Shakespeare speaks of death when Hamlet talks about the undiscovered country. In *Star Trek VI*, the undiscovered country is the future. You can think of Cloud NAS as both — the death of most on-premises infrastructure, and what the future holds in terms of global computing.

Don't wallow in the past. Don't mourn the shuttering of your data centers. Instead, look to the east. Feel the warmth of the rising sun on a new day. Free your mind from legacy constraints. Instead of thinking, "I wonder what we can do in the cloud," start saying "Why can't we do this in the cloud?"

Friends, the future is bright. The time is now.

Chapter 6

Ten Benefits of Hybrid Cloud NAS

In This Chapter

- Exploring how Avere Hybrid Cloud NAS can catapult you into the future

This chapter highlights how a Hybrid Cloud NAS solution from Avere Systems helps you cloud-enable your data center and lets you tackle the daily problems of scaling application performance and managing storage growth. It also helps you address long-term issues of delivering your product or service to market quickly and trying to keep up with explosive data growth and the associated spiraling costs.

- **Focusing on what's important:** Your business uses technology as a tool to achieve its strategic goals. To get ahead, you need to have laser-like focus on what makes you unique. Hybrid Cloud NAS can help you avoid wasteful reinventing of the wheel, letting you dedicate more of your resources to delivering value. You can burst storage and compute, or just go whole hog into the cloud! This newfound agility helps shorten cycle times, letting you get your products to market faster.
- **Maximizing your investment:** Hybrid Cloud NAS augments and extends the capabilities of your existing network attached storage (NAS) environments so that you're not replacing any gear you've already purchased or are still paying for. There's no need to get out of your local environment before your infrastructure reaches the end of its lifecycle.
- **Simplifying your architecture:** The Avere built-for-cloud file system enables deploying familiar NAS

infrastructures in the not-so-familiar cloud. Even if you are going cloud-native, you still have NAS as the building block of your architecture. Having a consistent view of storage, regardless of where it is, makes your life easier.

- **Getting performance where you need it:** With Avere, both physical and virtual FXT Edge filers track usage patterns of your files and move the most active files closer to the clients using them. With algorithms that keep the data close to your users, latency to clients is decreased, and the burden on your NAS and object-based Core filers is reduced.
- **Keeping pace with growing demand:** Demand for more application performance is growing every year, and this places an increasing demand on your storage. Avere Edge filers support clustering from 3 to 50 nodes to help you get there.
- **Storing data where it works best and costs the least:** Because the Edge-Core architecture of the Avere Hybrid Cloud NAS is aware of multiple types of storage media, each with its own performance and cost characteristics, your data is always placed where it makes the most sense, technologically and economically — whether that's on-premises or in the cloud.
- **Making sure that your data is always available:** Avere Edge clusters support high availability (HA) software that ensures that your data is always available even in the presence of network, hardware, and other failures.
- **Managing your data as a single pool:** Storage silos be gone! Avere Global Namespace (GNS) integrates all your Core filers, including NAS, private object, and public cloud, into a single pool of storage with transparent migration between them.
- **Consolidating and centralizing management:** By gathering together your NAS environments under an Avere Edge filer umbrella, you get a consolidated view of the performance of the entire NAS infrastructure and awesome storage analytics. With a single pane of glass for storage management, you don't have to deal with many individual management consoles.
- **Understanding usage and data flow better:** The ability to view what data is being accessed and with what regularity greatly increases your understanding of what your active and inactive storage areas are. This information helps you plan future purchases with more insight into where to focus your storage spending.

WHO USES AVERE

20 Top-grossing Blockbusters used Avere for special effects

2 of the top **3** Broadcast Companies

2 of the Forbes top **5** Biotech Companies for genomic sequencing

+ Top Hospitals

8 out of the largest **9** Technology Companies

run mission-critical applications on Avere or are Avere partners

Large Government Agencies

Including CDC, Library of Congress, NASA

Leading Hedge Funds use Avere for analytics and modeling in the cloud

3 out of **6** Super Major Oil & Gas Companies in their quest of oil

South America's **Largest Internet Service Provider**

More information at **averesystems.com**

Gain the benefits of the public cloud while maintaining on-premises storage performance

Competition is forcing companies to find innovative ways to apply compute and storage assets without expanding data centers. By leveraging the advantages of the public cloud, your company can increase storage and compute capacity — and save money at the same time. As on-premises equipment blends with what's available in the cloud, hybrid cloud can help you optimize your operations.

- ***Unlock the power of the cloud* — *join on-premises storage and compute resources with cloud storage and cloud compute***
- ***Change the center of gravity* — *move massive data sets from local NAS to efficient cloud providers***
- ***Burst compute to the cloud* — *add compute power when you need it, scale to the max, and pay only for what you use***
- ***Embrace the cloud* — *leave infrastructure to the giants so you can focus on your mission***

Open the book and find:

- **Cloud storage basics**
- **Differences between legacy computing and cloud computing**
- **Hybrid Cloud NAS case studies — genomics, media and entertainment, financial services, and government**
- **Ten benefits of Hybrid Cloud NAS**

Go to Dummies.com® for more!

Sharif Nijim is an IT architect and concurrent instructor of management at the University of Notre Dame. He tweets as @snijim on topics including APIs, cloud computing, devops, racing, and motorcycling.

FOR DUMMIES® A Wiley Brand

Also available as an e-book

ISBN: 978-1-119-36136-7
Not for resale

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.