



US 20250265156A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2025/0265156 A1

Rathi et al.

(43) Pub. Date: Aug. 21, 2025

(54) FAILOVER AND FAILBACK OF
DISTRIBUTED FILE SERVERS*G06F 16/11* (2019.01)
G06F 16/16 (2019.01)
G06F 16/182 (2019.01)
G06F 16/188 (2019.01)

(71) Applicant: NUTANIX, INC., SAN JOSE, CA (US)

(52) U.S. Cl.
CPC *G06F 11/1451* (2013.01); *G06F 11/1464* (2013.01); *G06F 11/2023* (2013.01); *G06F 11/203* (2013.01); *G06F 11/2092* (2013.01);
G06F 16/128 (2019.01); *G06F 16/168* (2019.01); *G06F 16/1844* (2019.01); *G06F 16/188* (2019.01); *G06F 2201/84* (2013.01)

(72) Inventors: Shyamsunder Prayagchand Rathi, Sunnyvale, CA (US); Hemanth Thummala, San Jose, CA (US); Lakshmana Reddy, San Jose, CA (US); Pradeep Thomas, San Jose, CA (US); Kalpesh Ashok Bafna, Fremont, CA (US); Manoj Premanand Naik, San Jose, CA (US)

(73) Assignee: NUTANIX, INC., SAN JOSE, CA (US)

(57) ABSTRACT

(21) Appl. No.: 18/970,664

(22) Filed: Dec. 5, 2024

Related U.S. Application Data

(63) Continuation of application No. 17/648,654, filed on Jan. 21, 2022, now Pat. No. 12,164,383.

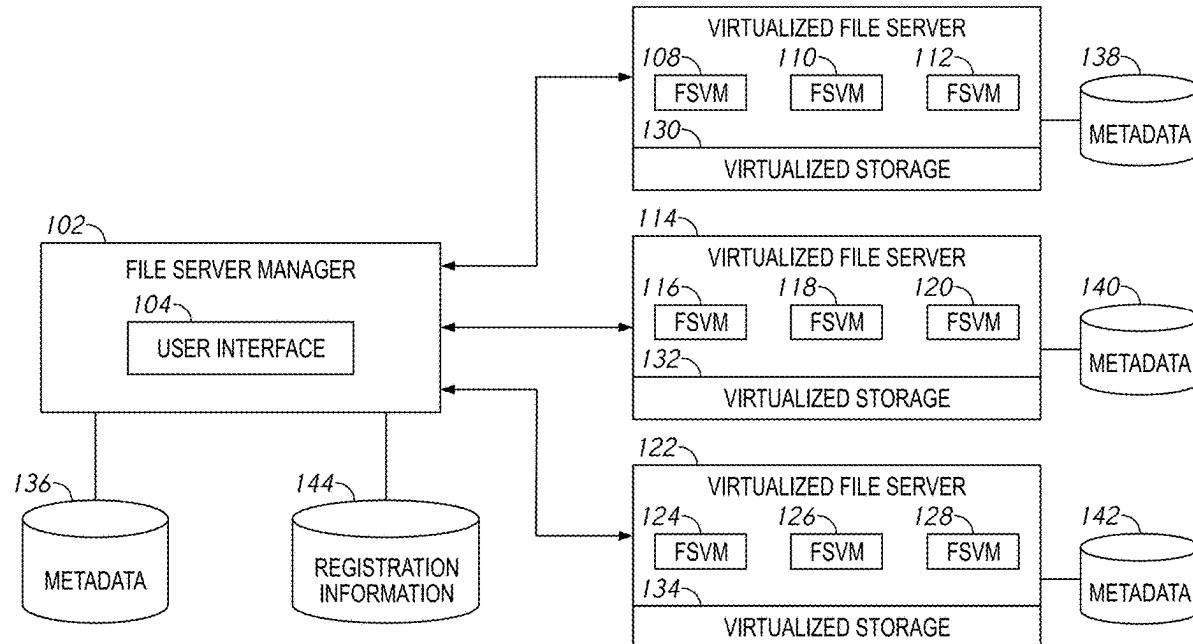
(60) Provisional application No. 63/260,438, filed on Aug. 19, 2021, provisional application No. 63/264,204, filed on Nov. 17, 2021.

Publication Classification

(51) Int. Cl.

G06F 11/14 (2006.01)
G06F 11/20 (2006.01)

An example file server manager updates a selected share of a destination distributed file server based on a snapshot of at least a portion of a selected share of a source distributed file server. The selected share of the destination distributed file server is updated while the source distributed file server serves client requests for storage items of the selected share of the source distributed file server. The file server manager receives a request to failover from the source distributed file server to the destination distributed file server and configures the destination distributed file server to service read and write requests for storage items of the selected share of the destination distributed file server. The file server manager further redirects client requests for storage items of the selected share of the source distributed file server to the destination distributed file server by updating active directory information.



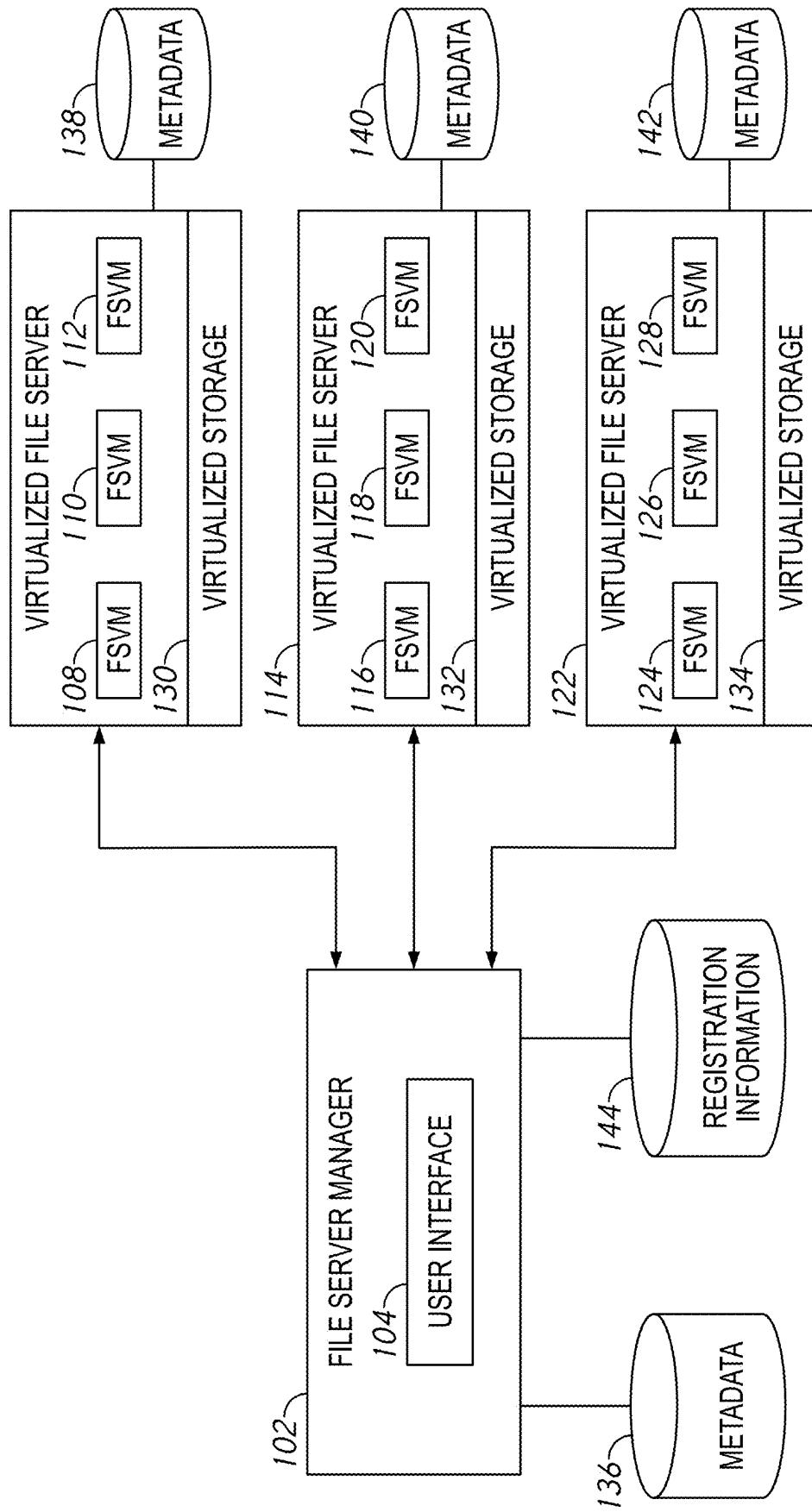


FIG. 1

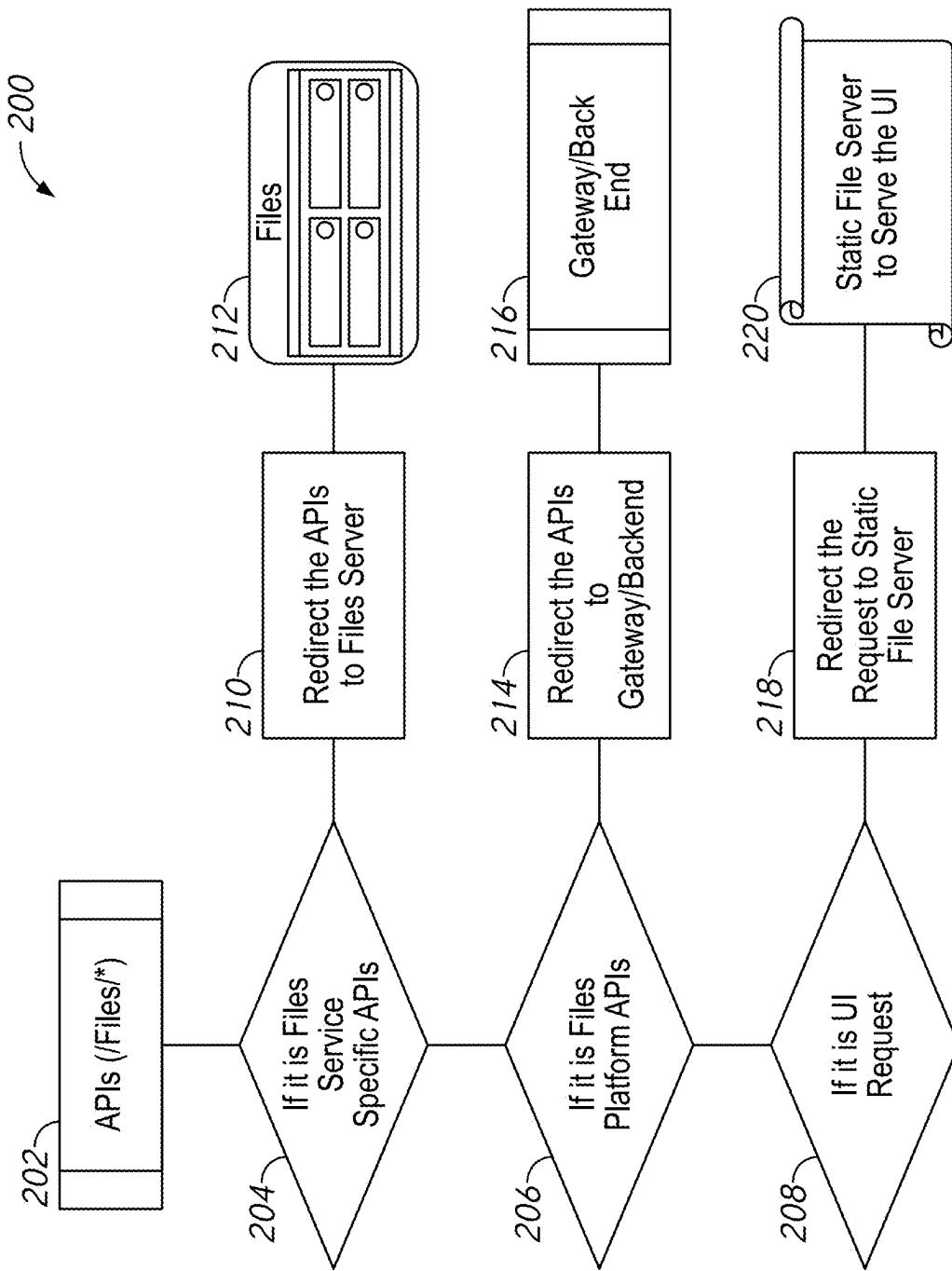


FIG. 2

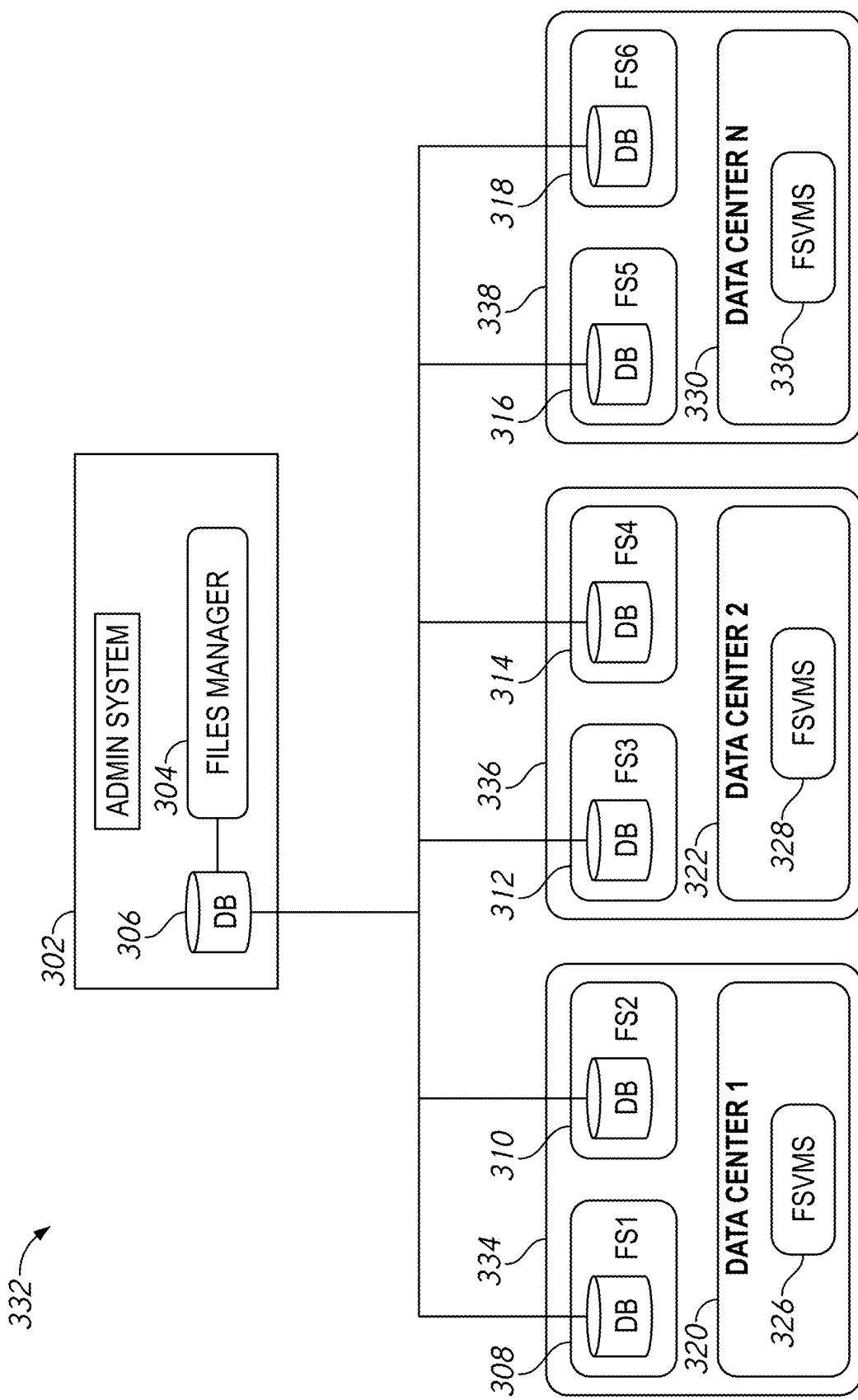


FIG. 3

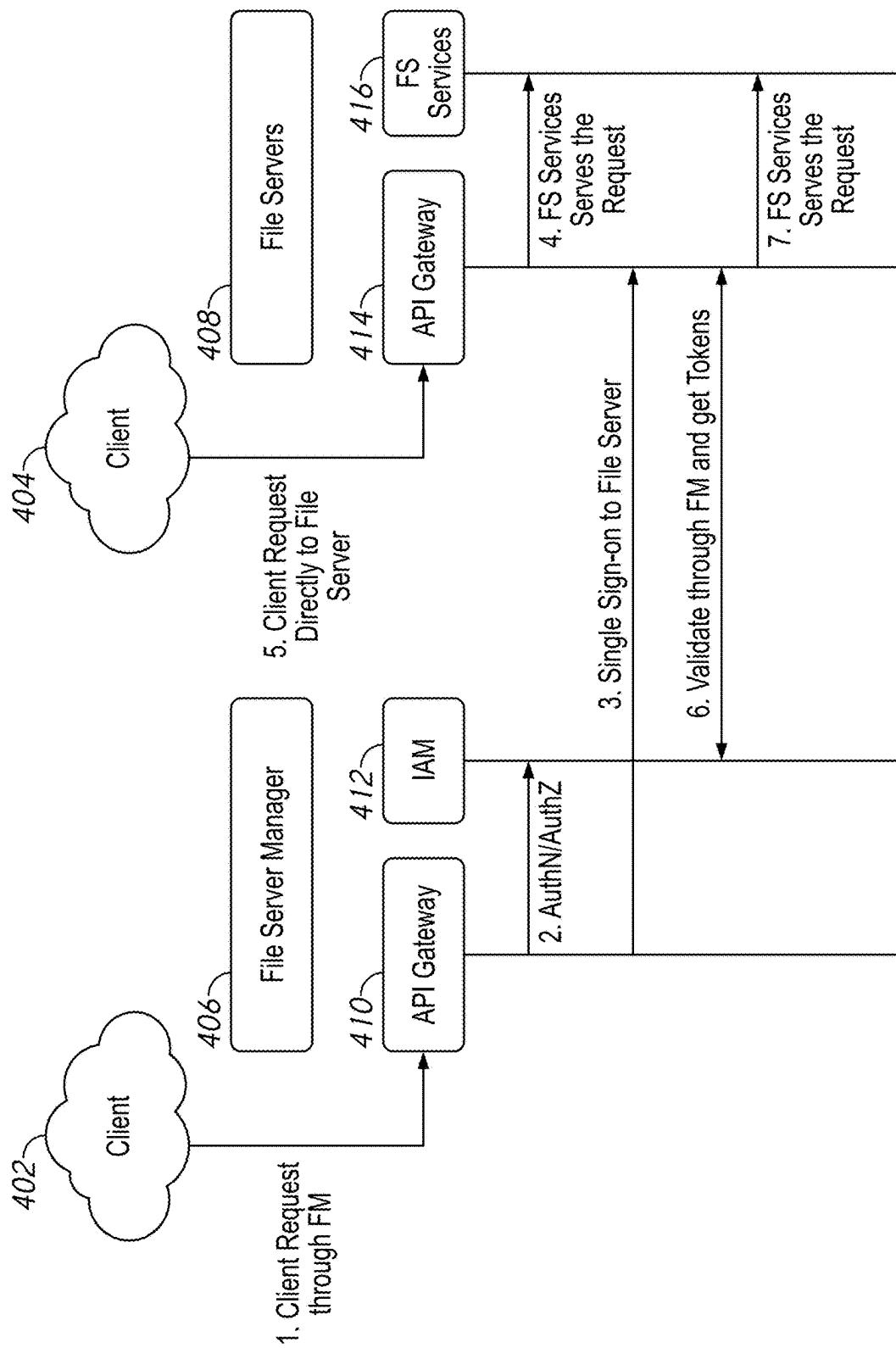


FIG. 4

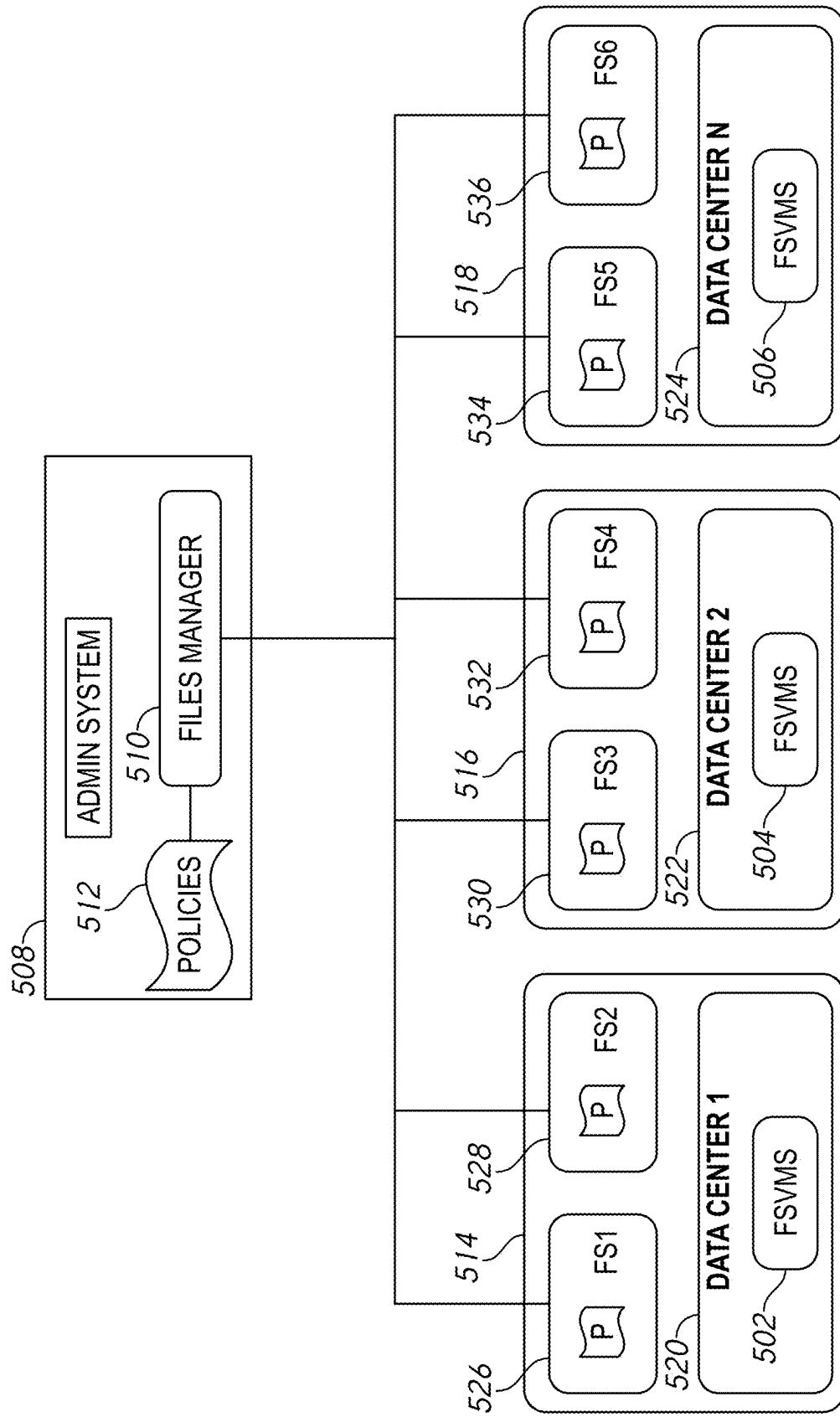


FIG. 5

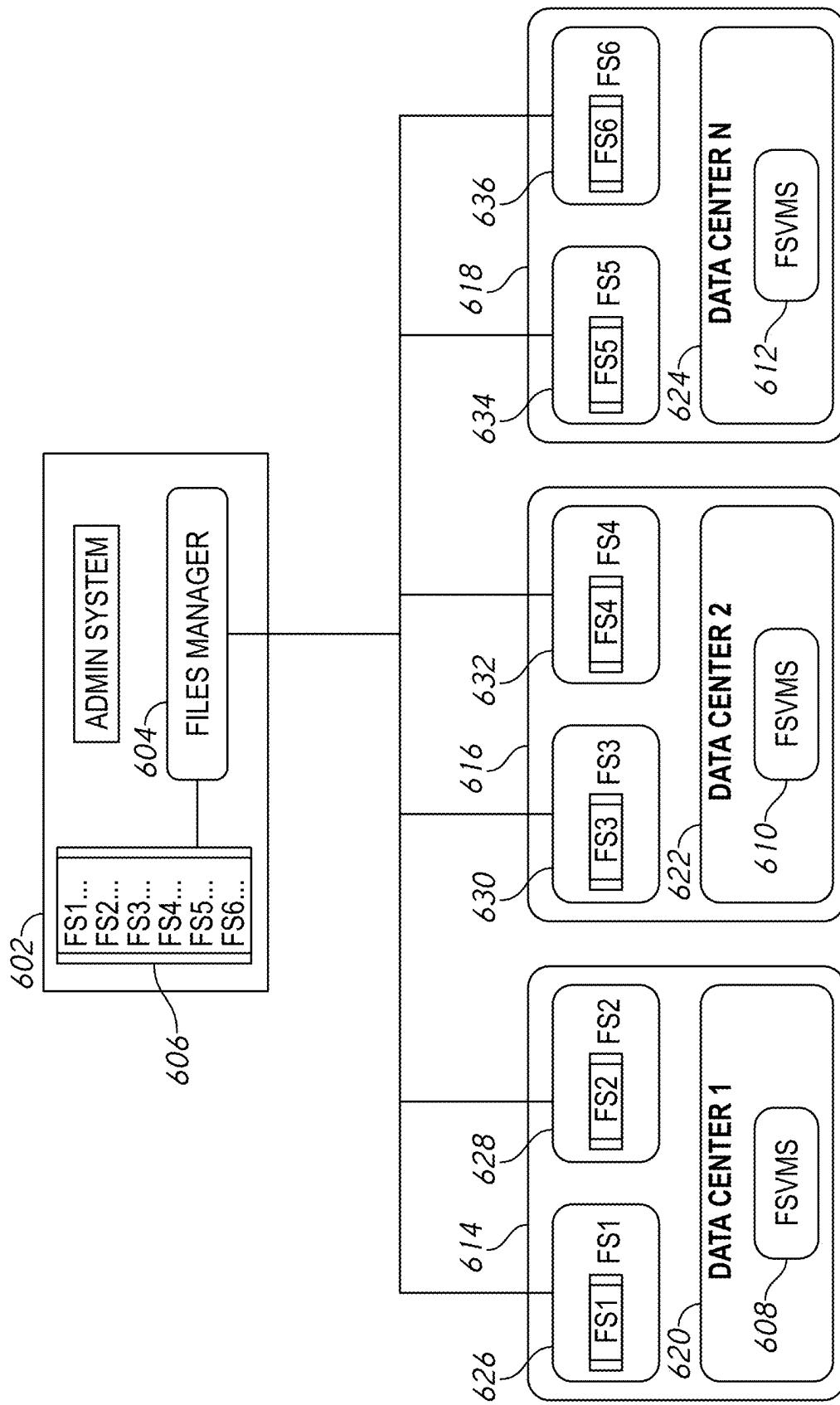


FIG. 6

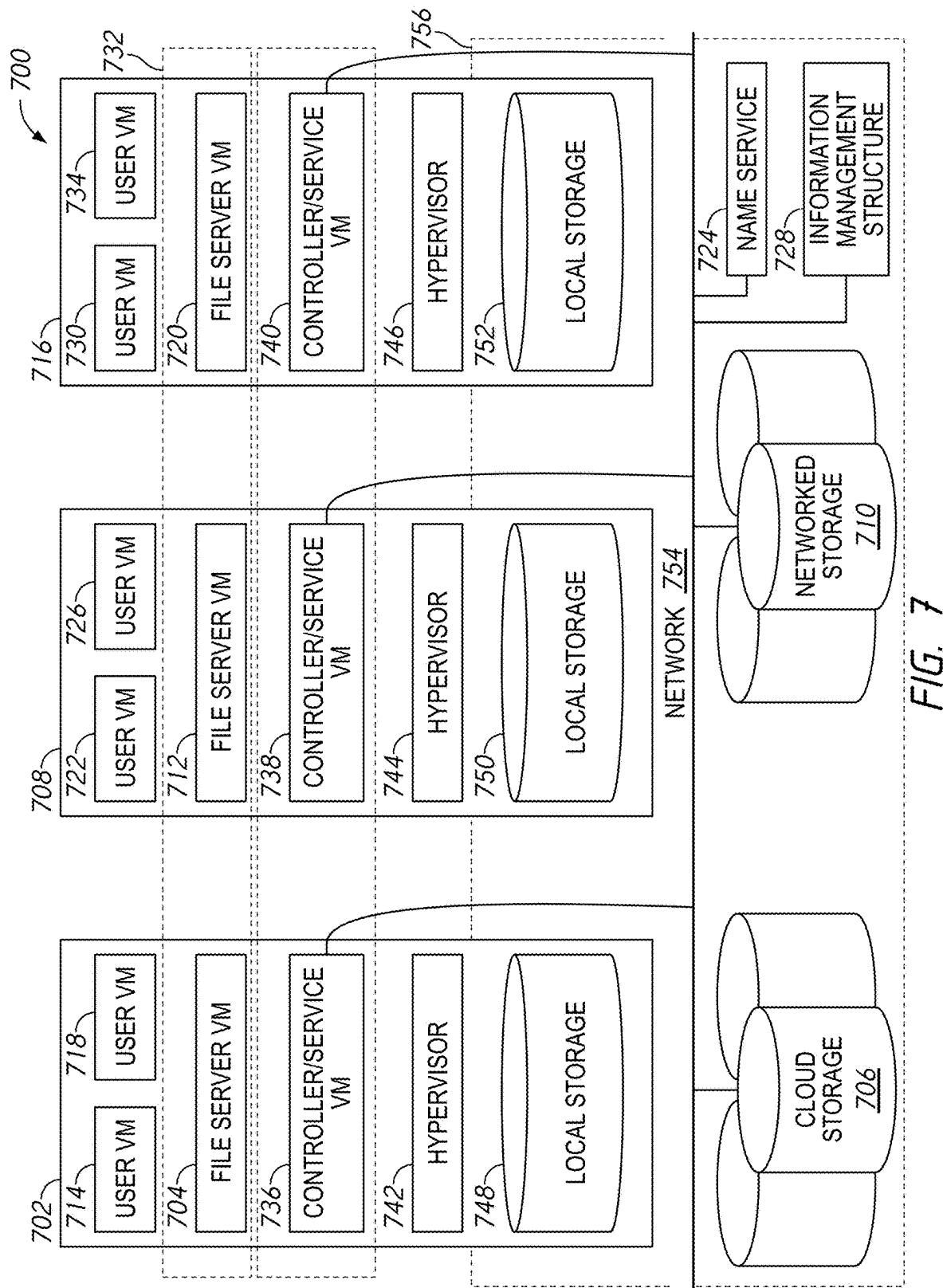
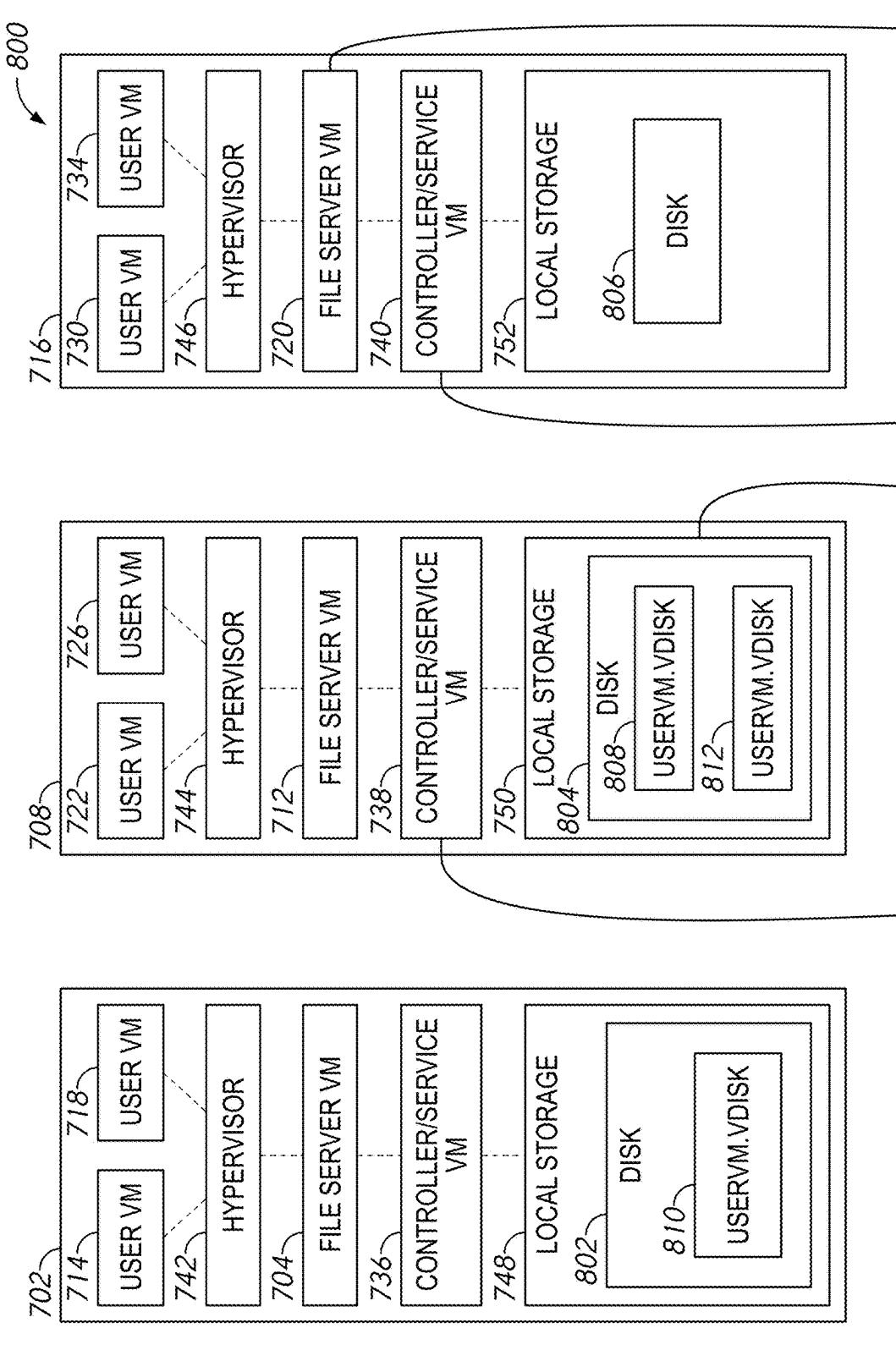


FIG. 7



NETWORK 754

FIG. 8

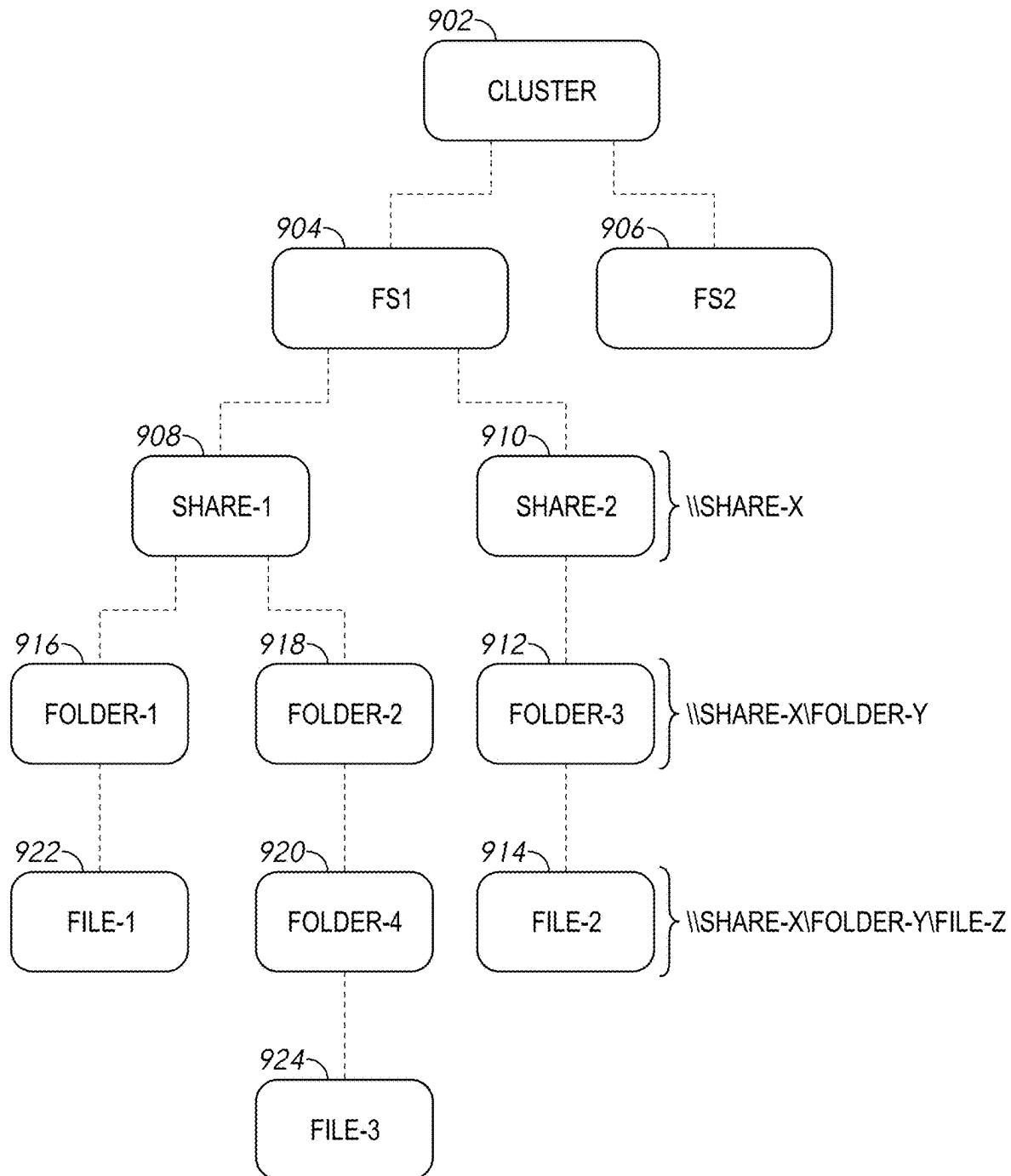


FIG. 9

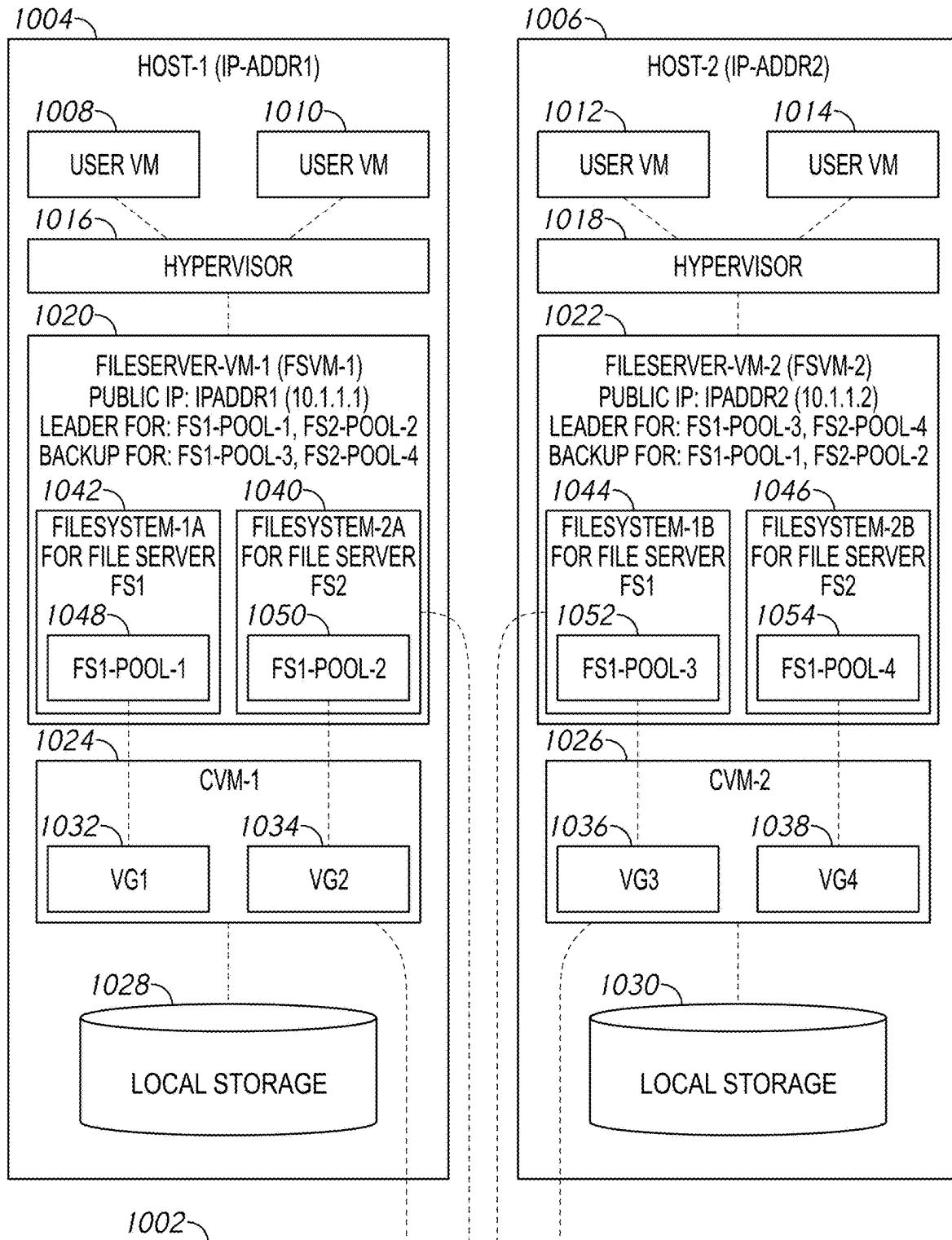


FIG. 10

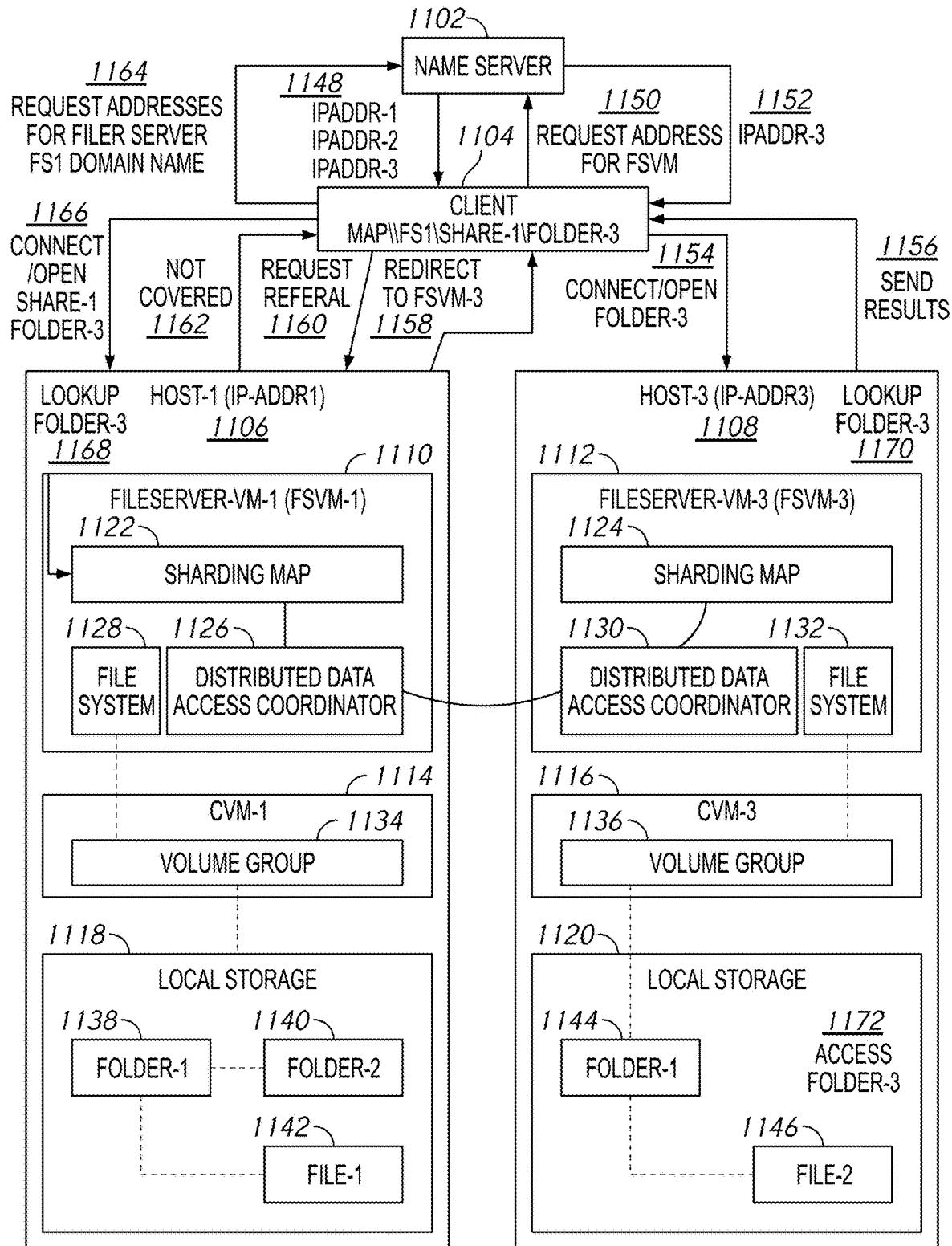


FIG. 11

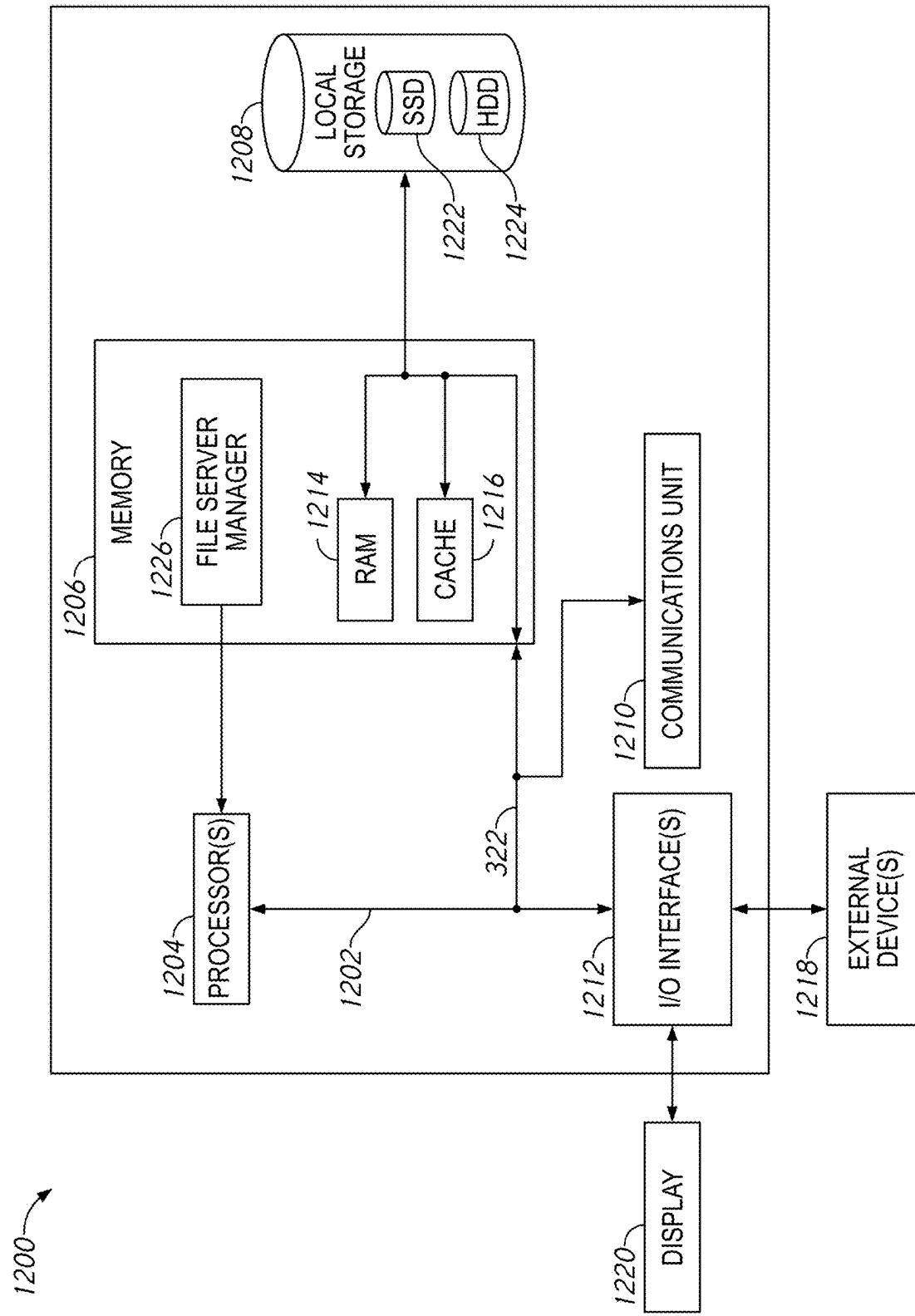


FIG. 12

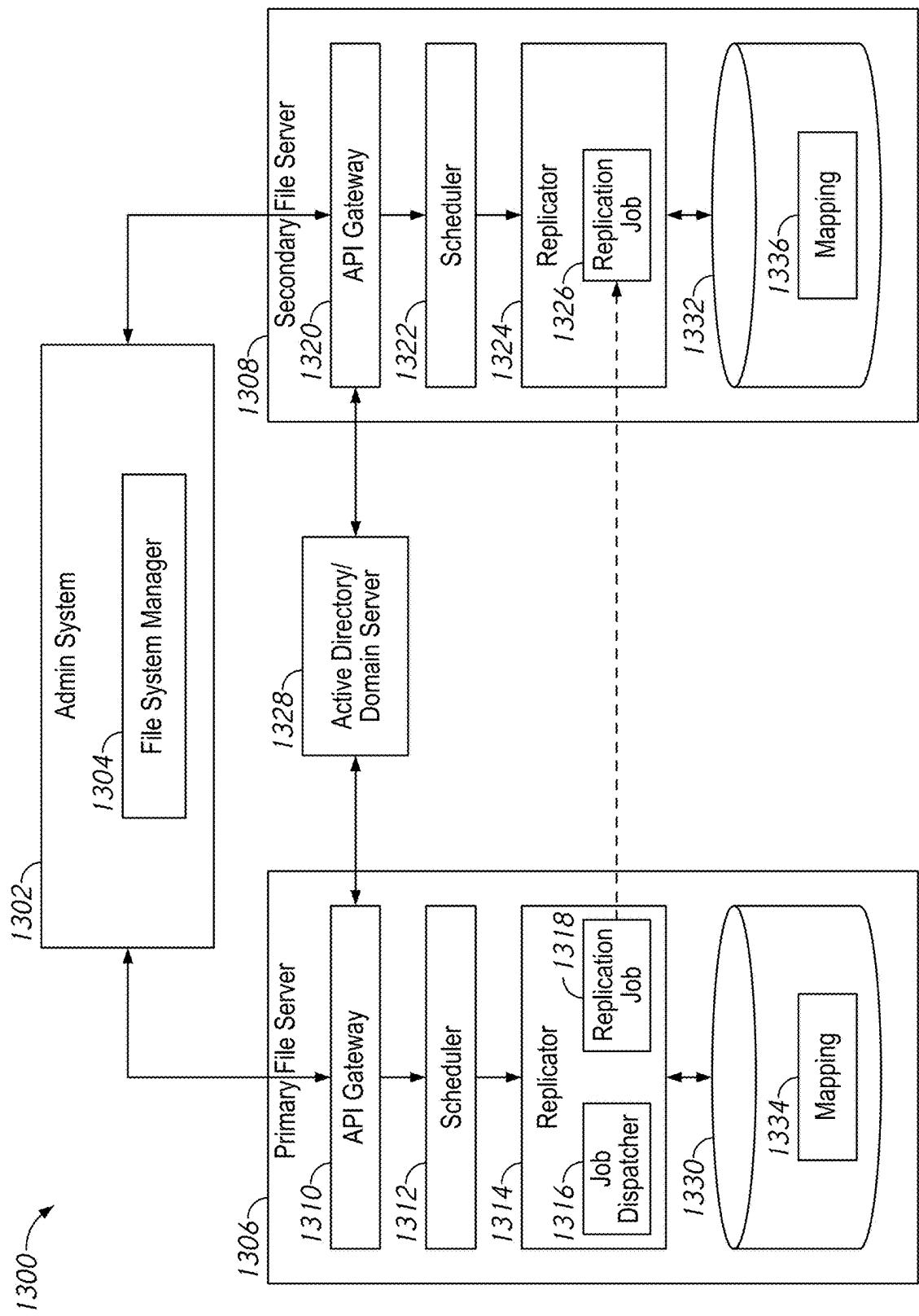


FIG. 13

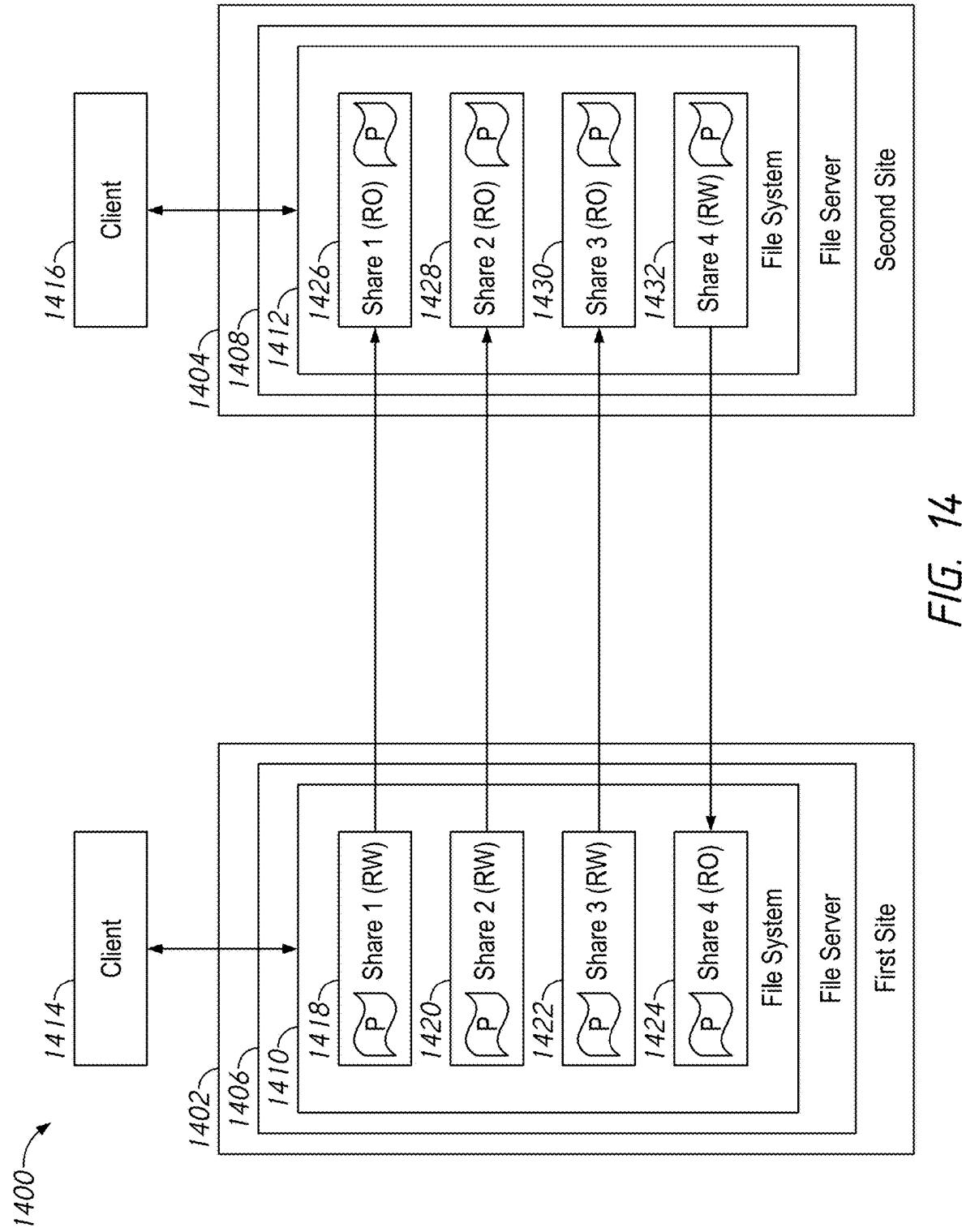


FIG. 14

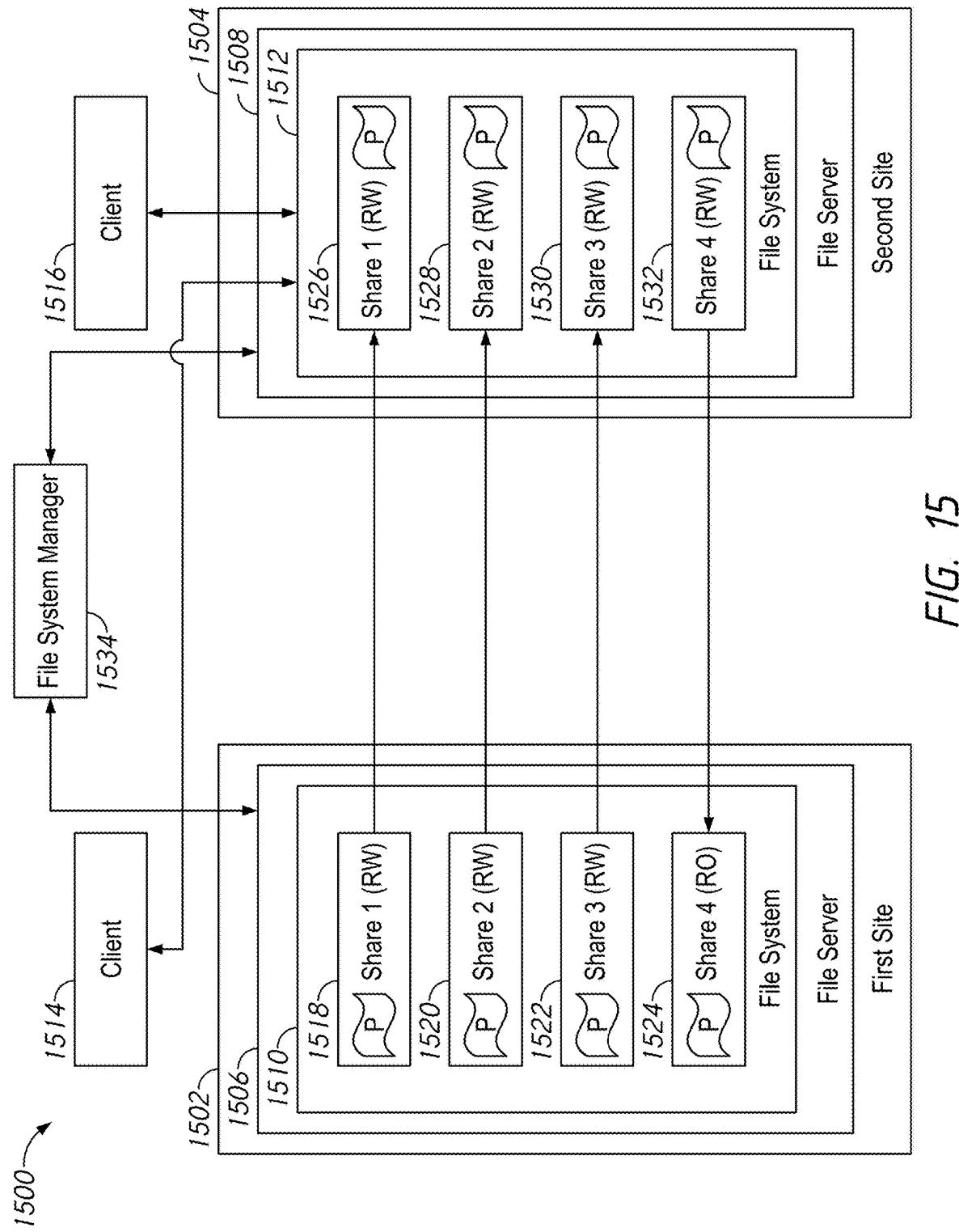


FIG. 15

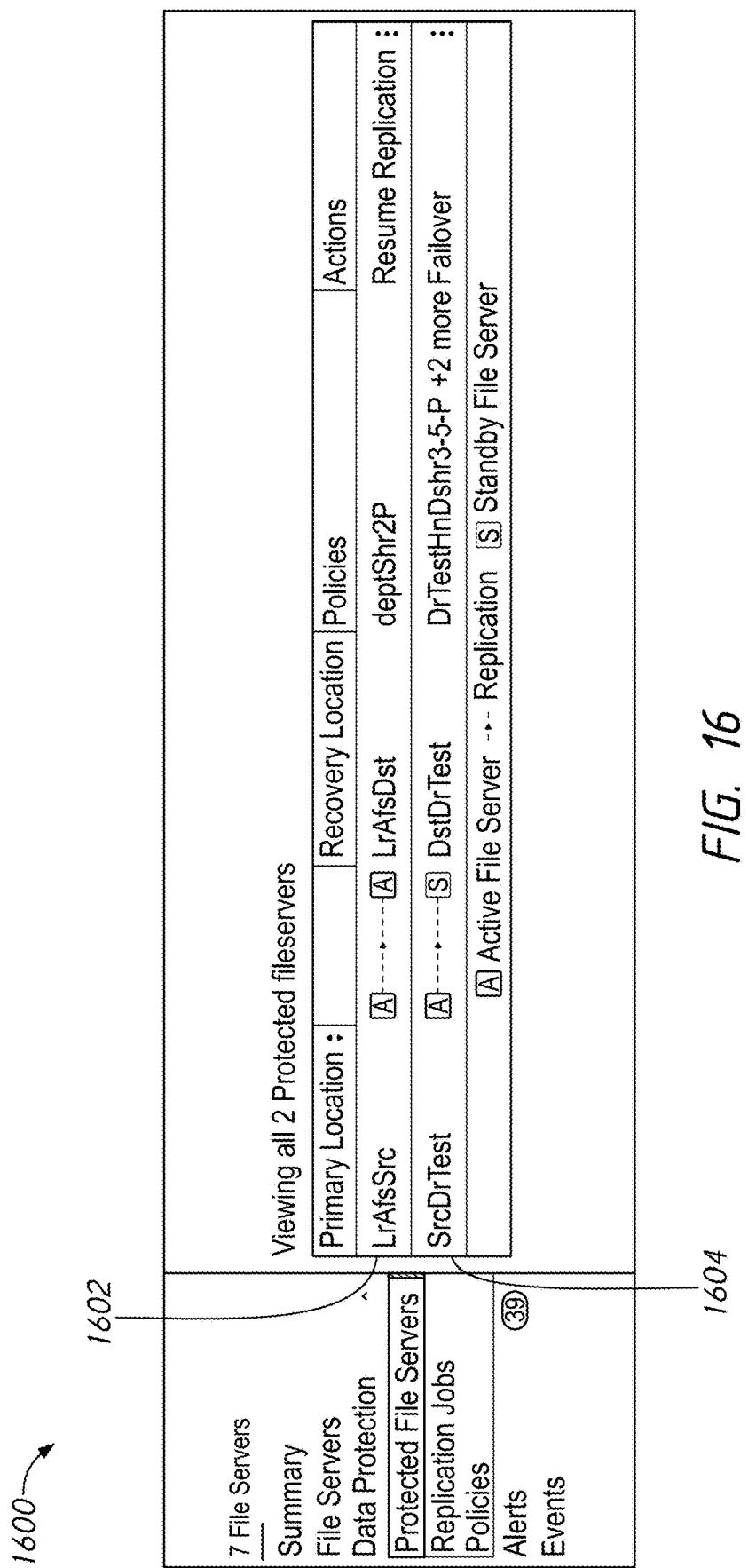


FIG. 16

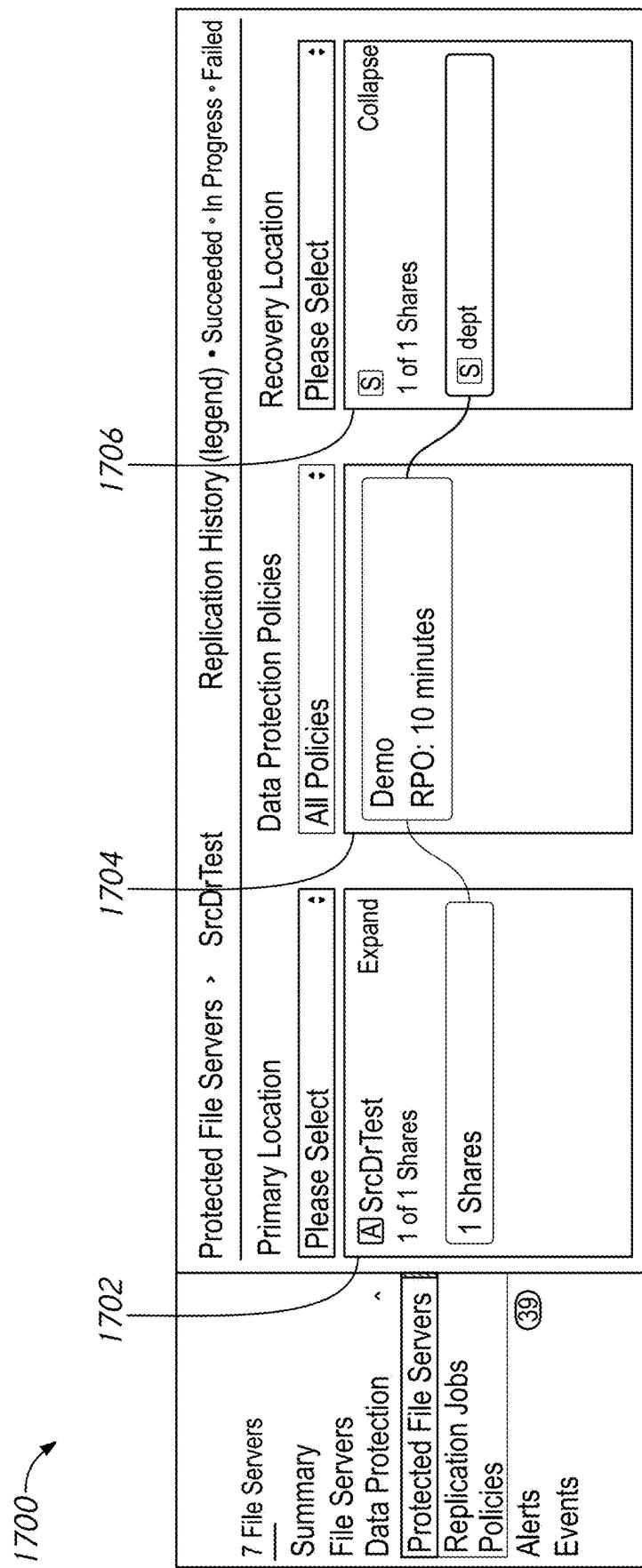


FIG. 17

1800 ↘

Viewing all 3,798 Replication Job

Start Time :	Duration :	RPO Compliance	Policy	Share	Source File Server	Target File Server	Filters
© 17:19:11	-	✓	regres	nsh1	regression	backup-380	JOB STATUS <input type="checkbox"/> Queued <input type="checkbox"/> In Progress <input type="checkbox"/> Succeeded <input type="checkbox"/> Cancelled <input type="checkbox"/> Failed
© 17:19:11	-	✓	regres	nnh1	regression	backup-380	POLICY <input type="checkbox"/> Contains <input type="text"/>
• 17:19:04	14 seconds	✓	regres	vdb15	regression	backup-380	SHARE <input type="checkbox"/> Contains <input type="text"/>
• 17:19:04	15 seconds	✓	regres	vdb16	regression	backup-380	SOURCE FILE SERVER <input type="checkbox"/> Contains <input type="text"/>
• 17:19:03	14 seconds	✓	regres	nsg1	regression	backup-380	
• 17:19:03	15 seconds	✓	regres	vdb14	regression	backup-380	
• 17:19:03	14 seconds	✓	regres	trump	regression	backup-380	
• 17:18:04	15 seconds	✓	smart	nm1-hom	multi-node-src	multi-node-dst	
• 17:18:03	15 seconds	✓	smart	nm1-hom	multi-node-src	multi-node-dst	
• 17:18:02	12 seconds	✓	smart	nm1-gen	multi-node-src	multi-node-dst	
© 1-10 of 3,798 © 110 rows ▾							

• Succeeded • Failed • Succeeded(RPO non-compliant)

FIG. 18

1900

Job Details	
State	Succeeded
Percentage Complete	100
Primary File Server	SrcDrTest
Recovery File Server	DstDrTest
Share	SharenoRO
Start Time	11:25:01
End Time	11:25:13
Duration	12s
Data synced	624 B
Network Throughput (avg)	52 B/sec

Close

FIG. 19

2000 ↗

7 File Servers		Actions			
	+ New Policy		ACTIONS		
Summary		Name	Primary Location	RPO	Recovery Location
File Servers		sharenoR0-P	SrcDr Test	Every 2 minutes	DstDr Test
Data Protection		testShr-P	SrcDr Test	Every 2 minutes	DstDr Test
Protected File Servers		testShr2P	LrAfssSic	Every 1 minutes	LrAfssDst
Replication Jobs		DrTestHrDsh3...	SrcDr Test	Every 5 minutes	DstDr Test
Policies		(39)			
Alerts					
Events					

FIG. 20

The screenshot shows a software interface for managing file replication policies. The main window title is "SharenoRO-P". On the left, there's a sidebar with links: "Back to policies", "Summary", "Shares", and "Replication Jobs". The main content area has two tabs: "Properties" and "Protection Status By Shares". The "Properties" tab displays the following configuration:

Description	Policy for sharenoRO
Source File Server	SrcDrTest
RPO	2 minutes
Recovery File Server	DstDrTest
Number of shares	1

The "Protection Status By Shares" tab shows one share with the status "RPO Compliant".

Annotations with callouts:

- An arrow labeled "2100" points to the "Back to policies" link.
- An arrow labeled "2102" points to the "Properties" tab.
- An arrow labeled "2104" points to the "Protection Status By Shares" tab.

FIG. 21

2200 ↘

↳ Back to policies

SharenoRO-P

Summary

Shares

Replication Jobs

Viewing 1 filtered shares

⌚ 1-1 of 1 ⌚ | 20 rows *

Share Name	Primary File Server	Recovery File Server	Last Successful Snapshot	Replication History
sharenoRO	SrcDr Test	DstDr Test	11:33:01	⌚ ⌚ ⌚ ⌚ ⌚ ⌚

* Succeeded • In Progress • Failed

FIG. 22

2300 ↗

Share Details		Viewing all 38 Replication Job						
Share Name	ShareNoRO	Start Time	Duration	RPD Compliance	Policy	Share	Primary File Server	Recovery File Server
Primary File Server	SrcDrTest	• 12 seconds	✓	share	share	SrcDrTest	DstDrTest	DstDrTest
RPO Compliance	• Compliant	• 12 seconds	✓	share	share	SrcDrTest	DstDrTest	DstDrTest
Last Successful Recovery Point	11:33:01	• 12 seconds	✓	share	share	SrcDrTest	DstDrTest	DstDrTest
		• 12 seconds	✓	share	share	SrcDrTest	DstDrTest	DstDrTest
		• 12 seconds	✓	share	share	SrcDrTest	DstDrTest	DstDrTest
		• 12 seconds	✓	share	share	SrcDrTest	DstDrTest	DstDrTest

Filters 1-20 of 38 20 rows

JOB STATUS

Queued 0
 Running 1
 Succeeded 33
 Cancelled 0
 Failed 5

PRIMARY FILE SERVER

Contains

FIG. 23

2400

Shares are marked active on both the file servers and are in Read/Write Mode.
Choose the direction in which you want to resume replication.

DstDrTest to SrcDrTest



SrcDrTest to DstDrTest



Shares on the SrcDrTest will be Active (Read/Write access) and the shares on the DstDrTest will be on Standby (Read only access)

Data Protection Policy

Policy Name
demo

Activated

Configure AD and DNS

For seamless experience for the clients or end users post resume_replication, the Active Directory and DNS entries of the recovery or standby server need to be updated with the primary or active server.

File Server Name

SrcDrTest

DNS Domain Name

child4.afs.minerva.com

Active Directory Credentials

Active Directory Realm Name

child4.afs.minerva.com

Username

Password

Show

Preferred Domain Controller

Optional

FIG. 24

2500 →

Resume Replication ×

Shares are marked active on both the file servers and are in Read/Write Mode. Choose the direction in which you want to resume replication.

DstDrTest to SrcDrTest SrcDrTest to DstDrTest

Shares on the DstDrTest will be Active (Read/Write access) and the shares on the SrcDrTest will be on Standby (Read only access)

Data Protection Policy

The shares on the Recovery are not protected by any policy. It is recommended to create a reverse protection policy between the Recovery and Primary which would help in easy failback process from recovery to primary once the primary is operational. All the shares from the primary server which are active on the recovery would be automatically added to this policy.

Create a Data Protection Policy
This policy would be created once you proceed with the resume action.

Recovery Point Objective (RPO)

1 Minutes

Policy Name

Reverse-demo

Close **Resume Replication**

FIG. 25

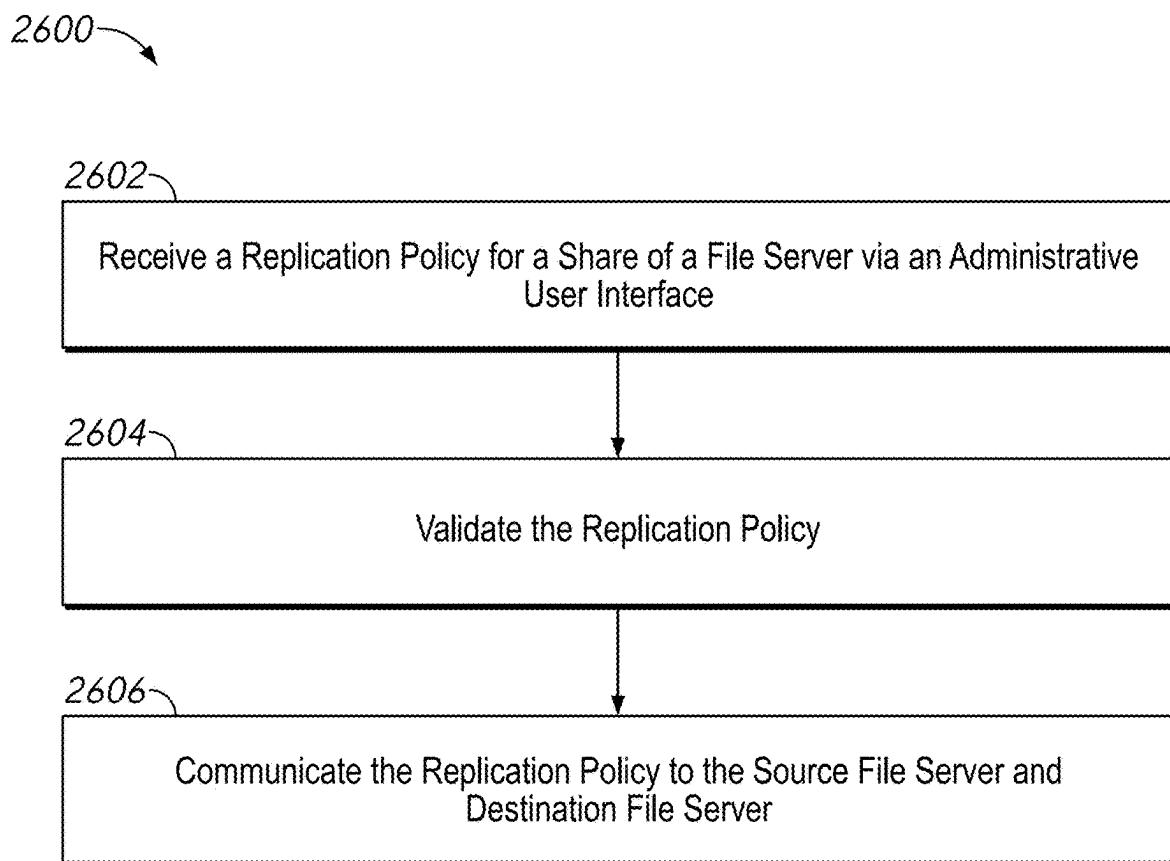


FIG. 26

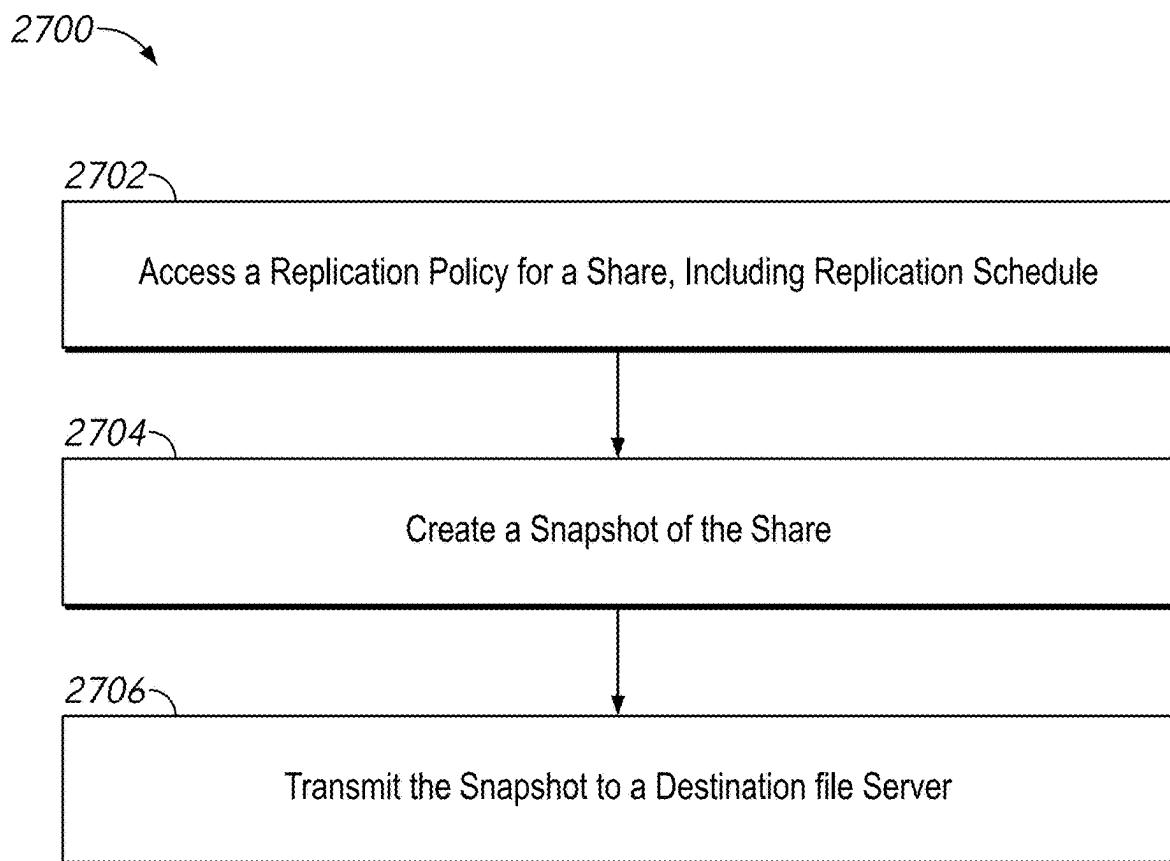


FIG. 27

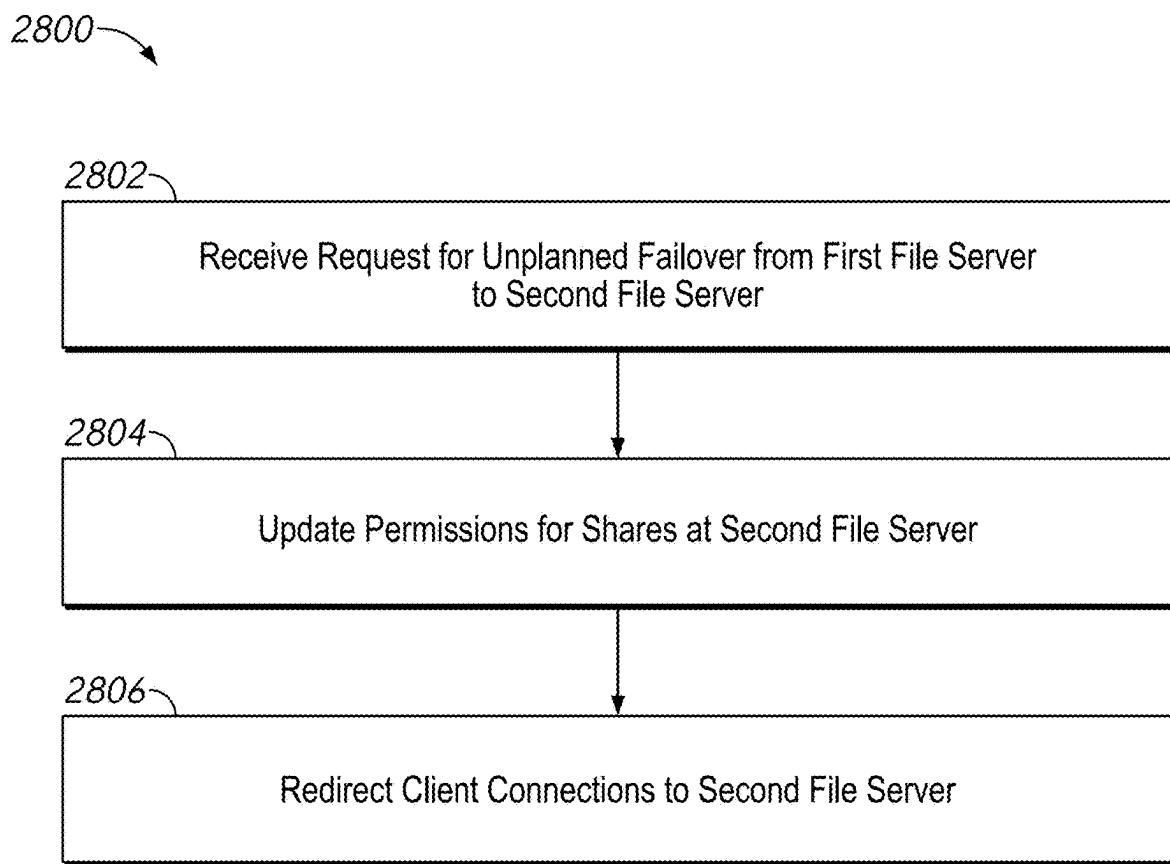


FIG. 28

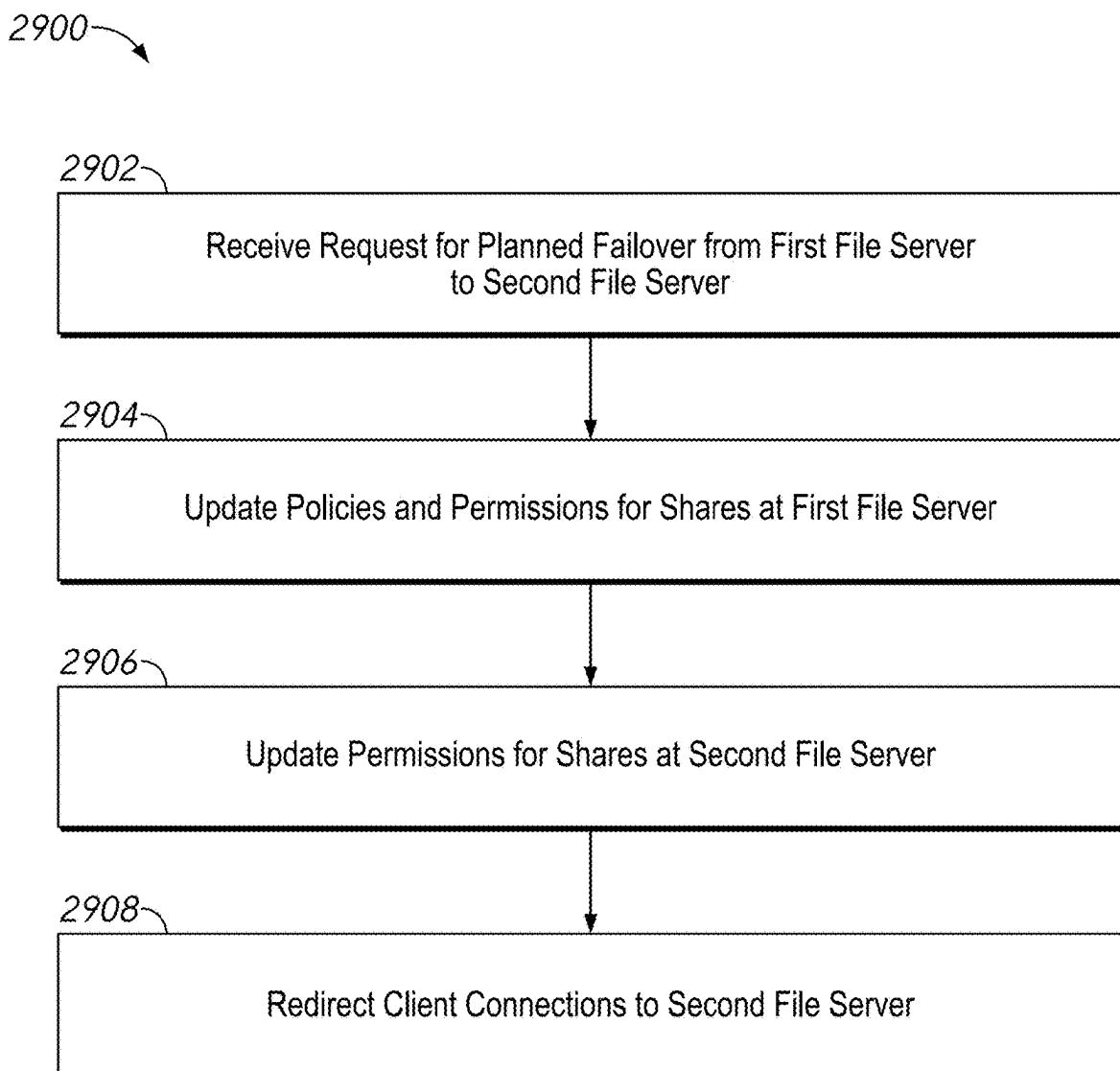


FIG. 29

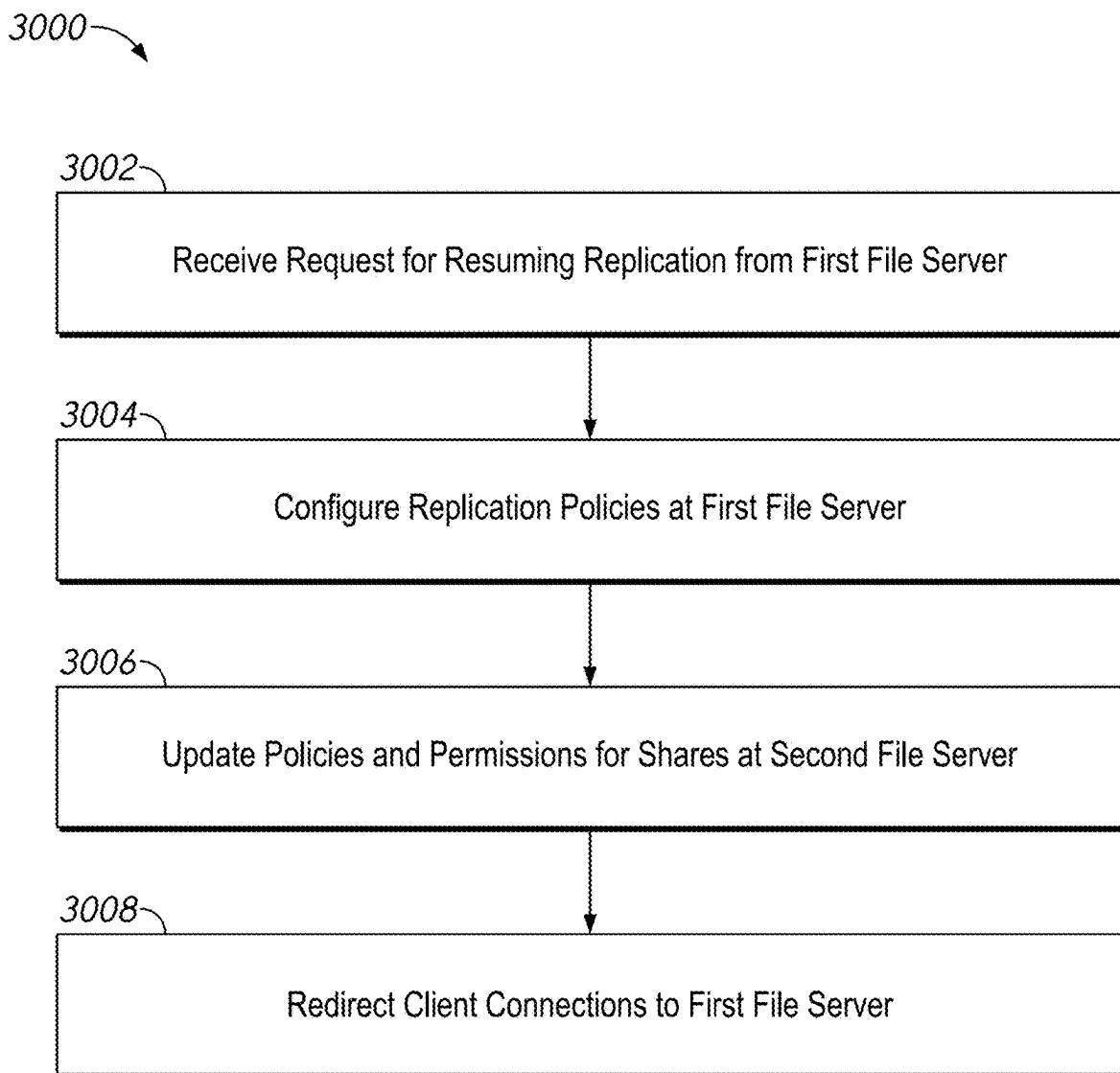


FIG. 30

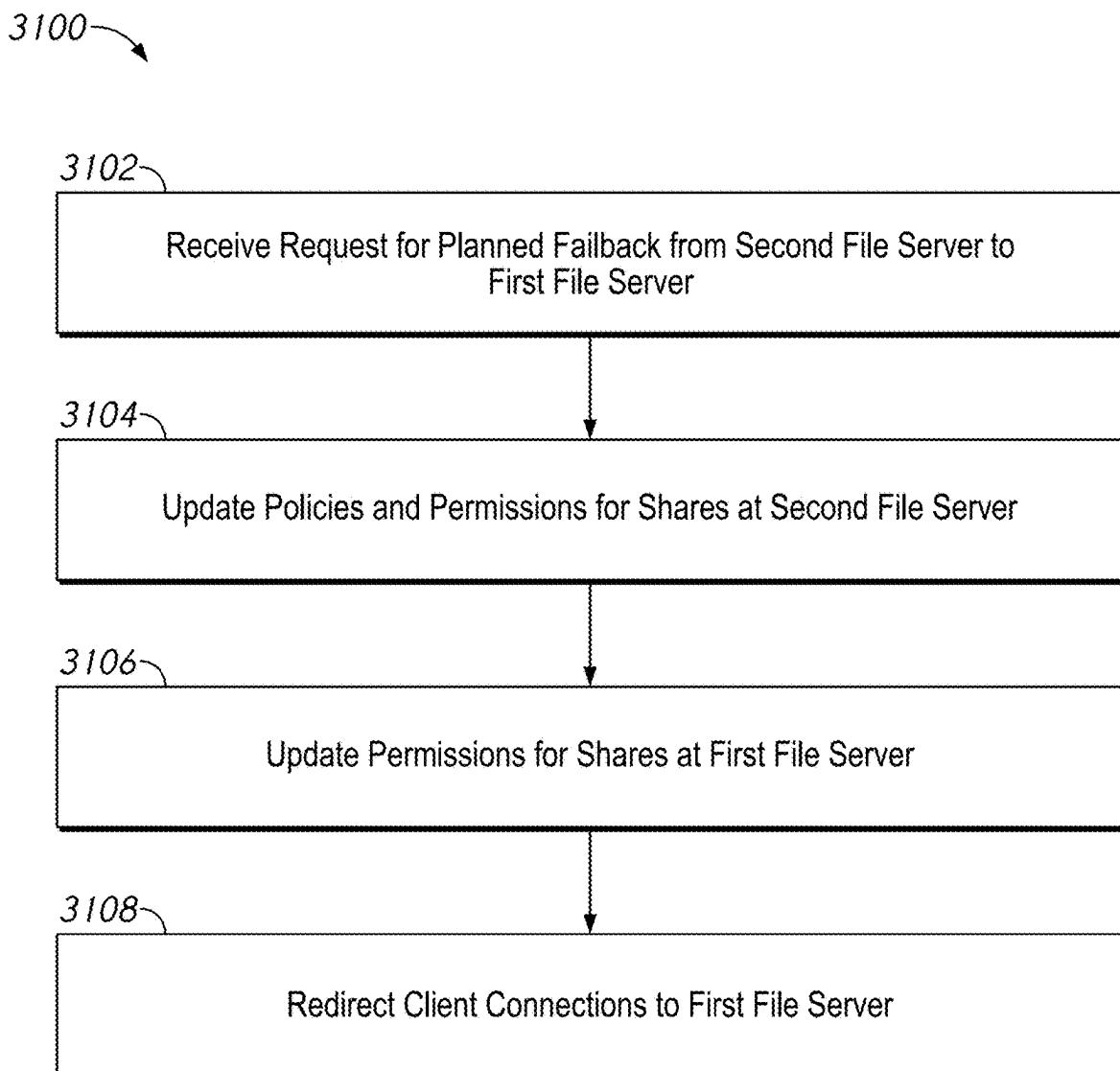


FIG. 31

FAILOVER AND FAILBACK OF DISTRIBUTED FILE SERVERS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. application Ser. No. 17/648,654 filed Jan. 21, 2022, which issues as U.S. Pat. No. 12,164,383 on Dec. 10, 2024, which claims priority to U.S. Provisional Application No. 63/260,438, entitled “File Server Managers and Systems for Managing Virtualized File Servers,” filed Aug. 19, 2021. This application claims priority to U.S. Provisional Application No. 63/264,204, entitled “Share-Based Disaster Recovery for Distributed File Server,” filed Nov. 17, 2021. The aforementioned applications are incorporated herein by reference, in their entirety, for any purpose.

BACKGROUND

[0002] Disaster recovery for distributed file servers often includes obtaining snapshots of a distributed file server. When a disaster occurs, failover (e.g., to a backup site) may take hours or longer to configure the file server at the backup site from the snapshots.

SUMMARY

[0003] An example computer readable medium disclosed herein is encoded with instructions which, when executed, cause a system to perform operations including updating a selected share of a destination distributed file server based on a snapshot of at least a portion of a selected share of a source distributed file server, where the selected share of the destination distributed corresponds to the selected share of the source distributed file server, where the selected share of the destination distributed file server is updated while the source distributed file server serves client requests for storage items of the selected share of the source distributed file server. The operations further include receiving a request to failover from the source distributed file server to the destination distributed file server and configuring the destination distributed file server to service read and write requests for storage items of the selected share of the destination distributed file server, where the configuring is responsive to the request to failover from the source distributed file server to the destination distributed file server. The operations further include redirecting client requests for storage items of the selected share of the source distributed file server to the destination distributed file server by updating active directory information for the source distributed file server to reference the destination distributed file server.

[0004] An example file server manager disclosed herein includes one or more processors and memory storing instructions which, when executed by the one or more processors, cause the file server manager to perform operations including updating a selected share of a destination distributed file server based on a snapshot of at least a portion of a selected share of a source distributed file server, where the selected share of the destination distributed corresponds to the selected share of the source distributed file server, where the selected share of the destination distributed file server is updated while the source distributed file server serves client requests for storage items of the selected share of the source distributed file server. The operations further include receiving a request to failover from the source

distributed file server to the destination distributed file server and configuring the destination distributed file server to service read and write requests for storage items of the selected share of the destination distributed file server, where the configuring is responsive to the request to failover from the source distributed file server to the destination distributed file server. The operations further include redirecting client requests for storage items of the selected share of the source distributed file server to the destination distributed file server by updating active directory information for the source distributed file server to reference the destination distributed file server.

[0005] An example method disclosed herein includes updating a selected share of a destination distributed file server based on a snapshot of at least a portion of a selected share of a source distributed file server, where the selected share of the destination distributed corresponds to the selected share of the source distributed file server, where the selected share of the destination distributed file server is updated while the source distributed file server serves client requests for storage items of the selected share of the source distributed file server. The method further includes receiving a request to failover from the source distributed file server to the destination distributed file server and configuring the destination distributed file server to service read and write requests for storage items of the selected share of the destination distributed file server, where the configuring is responsive to the request to failover from the source distributed file server to the destination distributed file server. The method further includes redirecting client requests for storage items of the selected share of the source distributed file server to the destination distributed file server by updating active directory information for the source distributed file server to reference the destination distributed file server.

[0006] Additional embodiments and features are set forth in part in the description that follows, and will become apparent to those skilled in the art upon examination of the specification and may be learned by the practice of the disclosed subject matter. A further understanding of the nature and advantages of the present disclosure may be realized by reference to the remaining portions of the specification and the drawings, which form a part of this disclosure. One of skill in the art will understand that each of the various aspects and features of the disclosure may advantageously be used separately in some instances, or in combination with other aspects and features of the disclosure in other instances.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The description will be more fully understood with reference to the following figures in which components are not drawn to scale, which are presented as various examples of the present disclosure and should not be construed as a complete recitation of the scope of the disclosure, characterized in that:

[0008] FIG. 1 is a schematic illustration of a system arranged in accordance with examples described herein.

[0009] FIG. 2 is a flowchart depicting a method of processing requests received at a file server manager arranged in accordance with examples described herein.

[0010] FIG. 3 is a schematic illustration of a system arranged in accordance with examples described herein.

[0011] FIG. 4 is a schematic illustration of a security workflow arranged in accordance with examples described herein.

[0012] FIG. 5 is a schematic illustration of a system arranged in accordance with examples described herein.

[0013] FIG. 6 is a schematic illustration of a system arranged in accordance with examples described herein.

[0014] FIG. 7 is a schematic illustration of a clustered virtualization environment 700 implementing a virtualized file server in accordance with examples described herein.

[0015] FIG. 8 is a schematic illustration of a clustered virtualization environment 800 arranged in accordance with examples described herein.

[0016] FIG. 9 illustrates an example hierarchical structure of a VFS instance in a cluster according to particular embodiments.

[0017] FIG. 10 illustrates two example host machines, each providing file storage services for portions of two VFS instances FS1 and FS2 according to particular embodiments.

[0018] FIG. 11 illustrates example interactions between a client and host machines on which different portions of a VFS instance are stored according to particular embodiments.

[0019] FIG. 12 is a schematic illustration of a computing system arranged in accordance with examples described herein.

[0020] FIG. 13 is a schematic illustration of a system arranged in accordance with examples described herein.

[0021] FIG. 14 is a schematic illustration of a system during replication arranged in accordance with examples described herein.

[0022] FIG. 15 is a schematic illustration of a system after failover arranged in accordance with examples described herein.

[0023] FIG. 16 illustrates an example user interface showing fileservers configured for backup and failover in accordance with examples described herein.

[0024] FIG. 17 illustrates an example user interface showing configuration details of a fileservers configured for backup and failover in accordance with examples described herein.

[0025] FIG. 18 illustrates an example user interface showing status of replication jobs in accordance with examples described herein.

[0026] FIG. 19 illustrates an example user interface showing job details in accordance with examples described herein.

[0027] FIG. 20 illustrates an example user interface showing replication policies in accordance with examples described herein.

[0028] FIG. 21 illustrates an example user interface showing details of a replication policy in accordance with examples described herein.

[0029] FIG. 22 illustrates an example user interface showing replication history of a share in accordance with examples described herein.

[0030] FIG. 23 illustrates an example user interface showing replication job history for a share in accordance with examples described herein.

[0031] FIG. 24 illustrates an example user interface for configuring share based replication in accordance with examples described herein.

[0032] FIG. 25 illustrates an example user interface for resuming share based replication in accordance with examples described herein.

[0033] FIG. 26 is a flowchart depicting an example method of configuring a replication policy for a share of a file server in accordance with examples described herein.

[0034] FIG. 27 is a flowchart depicting an example method of replicating a share of a file server in accordance with examples described herein.

[0035] FIG. 28 is a flowchart depicting an example method for unplanned failover in accordance with examples described herein.

[0036] FIG. 29 is a flowchart depicting an example method for planned failover in accordance with examples described herein.

[0037] FIG. 30 is a flowchart depicting an example method for planned fallback in accordance with examples described herein.

[0038] FIG. 31 is a flowchart depicting an example method of resuming replication at a first file server in accordance with examples described herein.

DETAILED DESCRIPTION

[0039] Various embodiments of the present disclosure will be explained below in detail with reference to the accompanying drawings. Other embodiments may be utilized, and structural, logical and electrical changes may be made without departing from the scope of the present disclosure.

[0040] Certain details are set forth herein to provide an understanding of described embodiments of technology. However, other examples may be practiced without various of these particular details. In some instances, well-known computing system components, virtualization operations, and/or software operations have not been shown in detail in order to avoid unnecessarily obscuring the described embodiments. Other embodiments may be utilized, and other changes may be made, without departing from the spirit or scope of the subject matter presented here.

[0041] Examples described herein include file server managers which may manage multiple virtualized file servers. The multiple virtualized file servers managed by a file server manager may be hosted by multiple computing node clusters (e.g., in multiple virtualization environments). The file server manager may be in communication with each of the multiple virtualized file servers. In this manner, a file server manager may provide a single pane of glass management interface to help manage and orchestrate file platform and service specific operations from a single location (e.g., a single logon and/or single user interface). File server managers may accordingly implement policies and conduct other operations based on data from multiple virtualized file servers in communication with the file server manager.

[0042] In various examples described herein, file server managers may be used to configure and manage share based replication of one or more virtualized file servers. For example, the file server manager may be configurable to select shares from a source virtualized file server to be replicated to a destination virtualized file server. The file server manager may further include replication policies controlling various aspects of share based replication. In some examples, file server managers may implement active-active share based replication, such that a backed-up virtualized file server (e.g., a destination virtualized file server) may be accessed and utilized relatively quickly upon failure

of a primary virtualized file server (e.g., a source virtualized file server). The file server manager may also, in some embodiments, implement fail-back to a source virtualized file server after a fail-over to a destination file server. File server managers may accordingly implement share based replication and/or backup of a virtualized file server.

[0043] Various examples of share based replication and file server level disaster recovery may provide a short recovery time, allowing for continuous availability of data after fail over to a recovery site. For example, configuration of active directory and domain name system entries may facilitate client access redirection from a failed file server to a backup file server after fail over. In the event of a planned or unplanned loss of service at a file server, corresponding shares at a backup or standby file server may be activated, and client connection requests for the file server may be redirected to the backup file server by updating address information for the file server at an active directory server or a domain name server. Accordingly, failover may occur quickly and with few noticeable changes to clients connecting to the backup file server.

[0044] FIG. 1 is a schematic illustration of a system arranged in accordance with examples described herein. The system of FIG. 1 includes file server manager 102. The file server manager 102 may provide user interface 104. The file server manager 102 may be in communication with memory and/or storage for metadata 136 and registration information 144. The system of FIG. 1 further includes virtualized file server 106, virtualized file server 114, and virtualized file server 122. The virtualized file server 106, virtualized file server 114, and virtualized file server 122 may each be in communication with the file server manager 102 (e.g., over one or more networks). Each of the virtualized file server 106, virtualized file server 114, and virtualized file server 122 may be hosted in a same and/or different virtualization environment. Each of the virtualized file server 106, virtualized file server 114, and virtualized file server 122 may include a cluster of computing nodes hosting a cluster of file server virtual machines (FSVM). For example, the virtualized file server 106 includes FSVM 108, FSVM 110, and FSVM 112. The virtualized file server 114 includes FSVM 116, FSVM 118, and FSVM 120. The virtualized file server 122 includes FSVM 124, FSVM 126, and FSVM 128. Each of the virtualized file server 106, virtualized file server 114, and virtualized file server 122 may include virtualized storage. For example, the virtualized file server 106 may include virtualized storage 130, the virtualized file server 114 may include virtualized storage 132, and the virtualized file server 122 may include virtualized storage 134. Moreover, each of the virtualized file server 106, virtualized file server 114, and virtualized file server 122 may include storage and/or memory for storing metadata. The virtualized file server 106 may store metadata 138. The virtualized file server 114 may store metadata 140. The virtualized file server 122 may store metadata 142.

[0045] The components shown in FIG. 1 are exemplary only. Additional, fewer, and/or different components may be used in other examples. For example, three virtualized file servers are depicted in FIG. 1, however any number may be used and may be in communication with the file server manager 102.

[0046] Examples of systems described herein may accordingly include one or more virtualized file servers, such as virtualized file server 106, virtualized file server 114, and

virtualized file server 122 in FIG. 1. A virtualized file server may represent a logical entity in the system. Virtualized file servers described herein may be hosted in generally any virtualization environment (e.g., on generally any virtualization platform). The virtualization environment and/or platform generally refers to the storage resources that have been virtualized by the virtualized file server and the compute resources (e.g., computing nodes with processor(s)) used to manage the virtualized storage. For example, the virtualized file server 106 may be hosted on a different virtualization environment than the virtualized file server 114 and/or than the virtualized file server 122. Nonetheless, in some examples one or more virtualized file servers in communication with a file server manager may be hosted in a same virtualization environment. Examples of virtualization environments include, for example, on premises installations of one or more computing nodes and storage devices. Examples of virtualization environment include one or more cloud computing systems (e.g., Amazon Web Services, MICROSOFT AZURE). Although not shown explicitly in FIG. 1, virtualization environments and/or virtualized file servers may include additional components including, but not limited to, one or more hypervisors, storage controllers, operating systems, and/or container orchestrators (e.g., Kubernetes). The multiple virtualized file servers in communication with a file server manager described herein may in some examples be located in different geographic locations (e.g., different buildings, states, cities, or countries).

[0047] A virtualized file server may include a cluster of virtual machines and/or other virtualized entities (e.g., containers), which may be referred to as file server virtual machines (FSVMs). In some examples, each of the file server virtual machines of a cluster may be implemented on different computing nodes forming a computing node cluster. For example, the FSVM 108, FSVM 110, and FSVM 112 of virtualized file server 106 may each be implemented on separate computing nodes of a computing node cluster used by the virtualized file server 106. Similarly, the FSVM 116, FSVM 118, and FSVM 120 may each be implemented on separate computing nodes of a computing node cluster used by the virtualized file server 114. Similarly, the FSVM 124, FSVM 126, and FSVM 128 may each be implemented on separate computing nodes of a computing nodes cluster. In some examples, a cluster of FSVMs may be implemented on a cloud computing system.

[0048] The FSVMs may operate to provide a file system on the storage resources of the virtualized file server. The file system may have a single namespace and may store data in accordance with filenames and/or directories. The FSVMs may accordingly support one or more file system protocols, such as NFS and/or SMB. A virtualized file server (such as virtualized file server 106, virtualized file server 114, and/or virtualized file server 122) may translate file system protocol requests for one or more files and/or directories (e.g., a file path) into one or more storage requests to access the data corresponding to the file, directory, and/or file path. Any of a variety of components of the virtualized file server may be used to perform the translation (e.g., one or more FSVMs, one or more hypervisors, and/or one or more storage controllers). The translation may be performed using a map (e.g., a shard map) relating the location of the data to the file name, share, directory, and/or file path.

[0049] Virtualized file servers described herein may include virtualized storage. For example, the virtualized file

server 106 may include virtualized storage 130. The virtualized file server 114 may include virtualized storage 132. The virtualized file server 122 may include virtualized storage 134. The virtualized storage may generally include any number or kind of storage devices—for example, network attached storage, local storage of one or more computing nodes forming the virtualized file server, and/or cloud storage. Storage devices may be implemented using, for example one or more memories, hard disk drives, solid state drives. The virtualized storage for a particular virtualized file server may be referred to as a storage pool. The virtualized storage may store one or more shares. Generally, the virtualized storage may refer to a storage pool which may include any of a variety of storage devices. In some examples, the virtualized file server(s) may be implemented in a hyper-converged architecture. For example, the storage pool may include local storage devices of the computing nodes used to host the virtualized file server. For example, virtualized storage 130 may include a storage pool. One or more shares of a file system provided by the virtualized file server 106 may be distributed across storage device of the storage pool, including local storage devices of one or more computing nodes on which the FSVM 108, FSVM 110, and/or FSVM 112 reside. In some examples, each file server virtual machine (FSVM) may manage (e.g., host) a corresponding share or portion of a share. A map may store associations between shares and files, directories, and/or file paths.

[0050] Virtualized file servers described herein may include metadata. For example, virtualized file server 106 may include metadata 138. The virtualized file server 114 may include metadata 140. The virtualized file server 122 may include metadata 142. The metadata may be stored, for example, in the virtualized storage and/or other storage location accessible to the virtualized file server. The metadata may in some examples be distributed across the storage pool of a virtualized file server. In some examples, the metadata may be stored in a database accessible to and/or hosted by the virtualized file server. Metadata stored by a virtualized file server may include, for example, authentication information for the virtualized file server and/or virtual machines in the virtualized file server, authorization information for the virtualized file server and/or virtual machines in the virtualized file server, configuration information for the virtualized file server and/or virtual machines in the virtualized file server, end point information (e.g., supported API calls and/or endpoints), a number of shares stored in the virtualized storage of the virtualized file server, a protocol supported by each share and/or FSVM (e.g., NFS and/or SMB), identities of the shares stored in the virtualized storage of the virtualized file server, a number of file server virtual machines (FSVMs) present in the virtualized file server, a number of files and/or directories hosted by the virtualized file server, compute resources available and/or used at the virtualized file server, storage resources available and/or used at the virtualized file server, or other metadata regarding the virtualized file server. The metadata may be maintained by the virtualized file server, for example, the metadata may be updated as the number of shares, FSVMs, storage resources and/or compute resources change.

[0051] Examples described herein may include a file server manager, such as file server manager 102 of FIG. 1. A file server manager may be in communication with multiple virtualized file servers. For example, the file server manager 102 may be in communication with virtualized file

server 106, virtualized file server 114, and virtualized file server 122. In this manner, the file server manager 102 may allow for access to, maintenance of, and/or management of multiple virtualized file servers (e.g., multiple file systems). An enterprise may have many virtualized file servers that are desired to be managed—for example, different geographic locations of the enterprise may maintain separate file systems and/or implement different privacy or other data policies. In some examples, different departments or entities within an organization may maintain respective virtualized file servers. An administrator or other entity associated with the enterprise, such as an IT manager, may advantageously view, access, and/or manage multiple virtualized file servers using the file server manager (e.g., file server manager 102). The file server manager may communicate with each virtualized file server using any of a variety of connections, including one or more networks. In some examples, a same network may be used to communicate between the file server manager and multiple virtualized file servers. In some examples, multiple networks may be used.

[0052] File server managers, such as file server manager 102 of FIG. 1 may be implemented using one or more computing devices. In some example, an administrative computing system may be used. The administrative computing system may include, for example, one or more processors and non-transitory computer readable media encoded with instructions for performing the file server manager operations described herein. In some examples, the file server manager may be implemented using a computing device different than the computing devices (e.g., computing nodes) used to implement the virtualized file server(s) with which the file server manager is in communication. In some examples, the file server manager may be hosted on one of the computing nodes forming a part of a virtualized file server in communication with the file server manager. File server managers, such as file server manager 102, may be hosted on premises systems in some examples, and/or on cloud computing systems in some examples.

[0053] Examples of file server managers described herein may provide one or more user interfaces, such as user interface 104 of FIG. 1. The user interface may allow a user (e.g., a human administrator and/or another computer process) to view information regarding multiple virtualized file servers, to communicate with multiple virtualized file servers, to manage multiple virtualized file servers, and generally to offer a single pane of glass interface to the multiple virtualized file servers in communication with the file server manager. The user interface may be implemented, for example, using one or more display(s) and one or more input and/or output device(s) (e.g., mouse, keyboard, touchscreen, etc.). In some examples, user interface 104 of file server manager 102 may be used to depict one or more of the virtualized file server 106, virtualized file server 114, and/or virtualized file server 122. For example, the identity and number of shares used by the virtualized file servers may be displayed. In some examples, the number and identity of computing nodes and/or FSVMs in each of the virtualized file servers may be displayed. Other attributes of the virtualized file servers may additionally or instead be displayed using a user interface of a file server manager. The data used in the display may wholly and/or partially be obtained from the registration information and/or metadata synchronized with one or more of the virtualized file servers.

[0054] Examples of file server managers described herein may store registration information, such as registration information 144 of FIG. 1. The registration information 144 may include information regarding each virtualized file server in communication with the file server manager. The registration information may include information used to manage, communicate with, and/or otherwise interact with the virtualized file server. Examples of registration information include a name of the virtualized file server, an identification of the virtualization environment hosting the virtualized file server, credentials for one or more FSVMs in the virtualized file server, IP addresses or other addresses for the virtualized file server, FSVMs in the virtualized file server, or other components of the virtualized file server. During setup of a system including a file server manager, the virtualized file servers may be registered with the file server manager, and may provide registration information to the file server manager. The registration information may be stored by the file server manager, such as in registration information 144, which may be a database in some examples. The registration information may be stored on a memory and/or other storage device accessible to the file server manager.

[0055] Examples of file server managers described herein may include metadata, such as metadata 136. The metadata may be synchronized to the metadata of multiple virtualized file servers in communication with the file server manager. For example, the metadata 136 may be synchronized with metadata 138, metadata 140, and metadata 142. For example, the metadata 136 at any given time may include metadata 138, metadata 140, and metadata 142. Synchronization may be maintained over time—the metadata of multiple virtualized file servers may periodically (e.g., at regular and/or irregular intervals) synchronize with the metadata store of the file server manager. In this manner, the file server manager 102 may maintain an updated storage of metadata associated with each of virtualized file server 106, virtualized file server 114, and virtualized file server 122. The metadata may be accessed by the file server manager and used to manage, communicate with, and/or otherwise interact with the virtualized file servers.

[0056] While the metadata 136 and registration information 144 are depicted separately in FIG. 1, they may be wholly and/or partially stored on a same storage device in some examples. The metadata 136 may be stored, for example, in a database. The registration information 144 may be stored, for example, in a database. Any of a variety of database synchronization techniques may be used to synchronize the metadata of the file server manager with the metadata of multiple virtualized file servers.

[0057] During operation, a file server manager described herein may register, such as by receiving a registration for, one or more virtualized file servers. For example, a virtualized file server (e.g., using an FSVM, a hypervisor, and/or another component of the virtualized file server), may transmit a registration (e.g., registration information) to the file server manager. In some examples, the file server manager may request such a registration by transmitting a request to register to the virtualized file server. In some examples, such as when the file server manager is hosted on a cluster and/or within a same system as the virtualized file server, an automatic registration may occur. For example, the registration process may include determining (e.g., from one or more IP addresses used), that a virtualized file server is

hosted on a same domain as a file server manager. In other examples, virtualized file servers which are not hosted on a same domain as a file server manager may nonetheless register with the file server manager. In the example of FIG. 1, the file server manager 102 may request registration from virtualized file server 106, virtualized file server 114, and virtualized file server 122. For example, a system administrator may enter an IP address, name, or other identifier to request a registration from virtualized file server 106, virtualized file server 114, and/or virtualized file server 122. In some examples, a system administrator or other user or component may transmit a registration from virtualized file server 106, virtualized file server 114, and/or virtualized file server 122, which registration may or may not be responsive to a request. In some examples, the operating system of one or more computing nodes of the virtualized file server hosting an FSVM may provide a registration request to the file server manager. The registration may include registration information which file server manager 102 may store in registration information 144.

[0058] The file server manager may synchronize metadata of registered file servers such that up to date metadata of the registered file server may be accessible to the file server manager. For example, the metadata 136 may synchronize with metadata 138, metadata 140, and metadata 142 of FIG. 1. Any and/or all types of metadata of the virtualized file server may be synced with a file server manager. For example, a number and identity of shares of each virtualized file server may be synchronized with the file server manager. In some examples, compute and/or storage resource usage may additionally or instead be synchronized between a virtualized file server and the file server manager. Sharding or other maps and/or portions thereof may be synchronized between a virtualized file server and the file server manager. Other metadata may be synchronized additionally or instead.

[0059] During operation, file server managers described herein, such as file server manager 102 of FIG. 1 may receive a management request for a particular virtualized file server. The management request may be received, for example by a client which may be hosted on a client system, on a system also hosting the file server manager, and/or on a system hosting all or a portion of one of the virtualized file servers in communication with the file server manager. In some examples, the management request may be implemented using an API call. In this manner, a file server manager may provide an API endpoint to receive API calls for one or more virtualized file servers. Examples of management requests include requests for accessing, managing, and/or maintaining the virtualized file server. For example, a management request may be a request to add and/or subtract one or more FSVMs, add and/or subtract one or more shares in the storage, and/or upgrade one or more FSVMs.

[0060] The file server manager may format the received management request for the virtualization environment (e.g., virtualization platform) used to host the requested virtualized file server. For example, the file server manager may access the registration information 144 to identify a virtualization environment for a virtualized file server identified in the management request. The management request may then be formatted in a manner used by the virtualized environment. In some examples, the formatted management request may be implemented as an API call, with the API call specific to the virtualization environment of the target vir-

tualized file server. In this manner, clients or other users providing management requests to the file server manager may not require knowledge of the virtualized environment hosting the virtualized file server. The file server manager may format the request in the manner used to communicate with the appropriate virtualization environment. This may provide flexibility in system design and usage, as multiple virtualization environments may be used, and virtualized file servers may in some examples be relocated from one virtualized environment to another without a need to update management requests being provided to the file server manager. Instead, an updated identification of the virtualized environment may be stored in registration information 144 and/or metadata 136.

[0061] During operation, the file server manager may utilize information from the registration to implement the management request. For example, access credentials provided during registration may be used to access one or more FSVMs and/or other components of the virtualized file server (e.g., hypervisor, other virtual machine(s) and/or container(s)) and implement the management request. In some examples, the management request may be provided to a particular FSVM. In some examples, the management request may be provided to an FSVM of the virtualized file server that is designated as a leader, and the leader FSVM may communicate the management request to an appropriate FSVM of the virtualized file server.

[0062] In some examples, file server managers described herein, such as file server manager 102 of FIG. 1, may be used to implement one or more cross-file server policies. A cross-file server policy may generally refer to a policy that accesses and/or utilizes more than one file server in implementing the policy. For example, one virtualized file server may be used (e.g., designated) as a destination file server and another virtualized file server may be used (e.g., designated) as a source file server. For example, the file server manager 102 may designate virtualized file server 106 as a source file server and virtualized file server 114 as a destination file server. The file server manager 102 may then utilize virtualized file server 106 to replicate, backup, provide redundancy for, or otherwise receive data from virtualized file server 106. For example, the file server manager 102 may implement a replication policy from virtualized file server 106 to virtualized file server 114. Without the presence of file server manager 102 in some examples, the virtualized file server 106 may have been used to implement a replication policy to virtualized file server 114 directly. However, utilizing file server manager 102 provides for central cross-server management and avoids a need for individual file servers to communicate with one another directly.

[0063] FIG. 2 is a flowchart depicting a method of processing requests received at a file server manager arranged in accordance with examples described herein. The flowchart 200 depicts management requests that may be provided to a file server manager as one or more API calls, such as API call 202. The flowchart 200 includes evaluating the type of management request—is it files service specific in block 204, is it files platform specific in block 206, is it a UI request in block 208. Responsive to a determination the management request is files service specific in block 204, the flowchart 200 indicates the API may be redirected to a virtualized file server in block 210, and may be provided, for example, to virtualized file server 212. Responsive to a determination the management request is files platform

specific in block 206, the request may be directed to a gateway or backend in block 214, such as gateway 216. Responsive to a determination the management request is a UI request, it may be redirected to a file server in block 218, such as file server 220, which may be a virtualized file server. The blocks and components of flowchart 200 are exemplary, and the blocks and component may occur in different orders in some examples, and additional and/or fewer blocks or components may be used in some examples.

[0064] The method depicted by flowchart 200 may be implemented by file server managers described herein, such as by file server manager 102 of FIG. 1.

[0065] In block 202, an API call 202 may be received, which may also be referred to as a management request. The API call 202 may be implemented, for example, using a REST API. The API call 202 may be received from an administrator (e.g., using an interface to a file server manager, such as user interface 104 of FIG. 1). The API call 202 may be received from a computing system (e.g., a client computing system) in communication with a file server manager described herein. In some examples the request may come from an automation script that may be executing on, for example a computing system in communication with the file server manager and/or on the file server manager itself. In some examples, the API call may be received from (e.g., may be sent by) a virtualized file server in communication with the file server manager, such as virtualized file servers 106, virtualized file server 114, and/or virtualized file server 122 of FIG. 1. The API call may be received from a virtual machine and/or container. For example, the API call may be received from a user virtual machine and/or container which may be hosted on a same computing node of as one of the FSVMs of the virtualized file server. The API call may be used to implement a management request as described herein. The API call may not be specific to the type of platform (e.g., virtualization platform) hosting a virtualized file server. In this manner, the API call may be agnostic to platform type. The file server manager may accordingly provide an API endpoint for management requests directed to one or more virtualized file servers.

[0066] The file server manager may evaluate the management request received, e.g., API call 202. For example, the file server manager 102 may receive API call 202 and may evaluate it to determine how to direct the API call. The management request may be evaluated to determine its intended destination. For example, the file server manager 102 may evaluate a management request to determine if it is files service specific (e.g., in block 204), if it is directed toward a files platform (e.g., block 206), and/or if it is a UI request (e.g., block 208). The evaluation may be based, for example, on identifying that the content of the request pertains to files services, files platform, and/or UI. The evaluation may be based, for example, on identifying a destination of the request.

[0067] If the management request (e.g., API call 202) is determined to be a files service specific request in block 204, the request may be redirected to the appropriate virtualized file server in block 210. Examples of files service specific requests include requests to create a share, create or revise one or more user quotas for the virtualized file server, monitor a number of users connected to a virtualized file server, or blocking one or more particular users of the virtualized file server. Files service specific requests may not need to be translated for the particular virtualization plat-

form of the virtualized file server, because in some examples they may receive requests which are received and/or processed by one or more file server virtual machines (FSVMs) or another component of the virtualized file server (e.g., hypervisor, daemon, or other service). In redirecting to a virtualized file server, the file server manager may in some examples format the request in a manner suitable for the virtualized file server, such as the virtualized file server 212. In some examples, the file server manager may format the request in a manner suitable for a particular version of file server virtual machine operating in the requested virtualized file server. In some examples, the file server manager may receive a request for a virtualized file server that the particular version of file server virtual machine used may not support. The file server manager may identify the version of file server virtual machine (e.g., by accessing metadata and/or registration information) and may replace the unsupported request with a supported request able to be received and processed by the version of file server virtual machine in operation.

[0068] If the management request (e.g., API call 202) is determined to be a files platform request (e.g., create one or more FSVMs, scale-in the virtualized file server, scale-out the virtualized file server, add storage to the virtualized file server), then management request may be redirected to a gateway or backend for the appropriate virtualized file server in block 214, such as gateway 216. In redirecting the request, the file server manager may format the request for the particular virtualization platform (e.g., virtualization environment) on which the virtualized file server is hosted. For example, the management request may be formatted for the compute and storage resources used in a particular environment such as a NUTANIX platform, an AMAZON WEB SERVICES platform, a MICROSOFT AZURE platform, etc. The file server manager may access a database or other stored location to determine the platform hosting the requested virtualized file server (e.g., registration information 144 in FIG. 1). Based on the identity of the platform, the management request may be formatted for the platform (e.g., by utilizing platform-specific API calls in some examples). In this manner, an administrator or other user may manipulate a virtualized file server—e.g., to expand the virtualized file server-without requiring knowledge on behalf of the administrator of what platform is hosting the virtualized file server.

[0069] If the management request (e.g., API call 202) is determined to be a UI request, it may be redirected to a file server in block 218, such as file server 220, which may be a virtualized file server. UI requests may include, for example, requests to view the current compute resource usage, storage resources usage, number of shares, identity of shares, and/or files or directories hosted by a particular virtualized file server. In redirecting the request to a file server in block 218, the file server manager may format the request in a manner particular to the file server and/or the UI of the file server.

[0070] Accordingly, using methods such as depicted in flowchart 200 of FIG. 2, file server managers described herein may receive and redirect management requests, such as API calls. The API calls may be selected and/or formatted in a manner particular to a virtualized file server and/or a virtualization environment.

[0071] FIG. 3 is a schematic illustration of a system arranged in accordance with examples described herein. The

system 332 includes an admin system 302 in communication with virtualized file server 334, virtualized file server 336, and virtualized file server 338. The virtualized file server 334 is hosted on virtualization platform 320. The virtualized file server 336 includes FSVMs 326 and is hosted on virtualization platform 322. The virtualized file server 336 provides file system 308 and file system 310. The virtualized file server 338 includes FSVMs 328 and is hosted on virtualization platform 324. The virtualized file server 338 provides file system 312 and file system 314. The virtualized file server 338 includes FSVMs 330 and is hosted on virtualization platform 324. The virtualized file server 338 provides file system 316 and file system 318. The admin system 302 includes file server manager 304 and database 306. The components shown in FIG. 3 are exemplary only. Additional, fewer, and/or different components may be used in other examples.

[0072] In some examples, the system 332 of FIG. 3 may be used to implement and/or may be implemented by the system of FIG. 1, or portions thereof. For example, the file server manager 304 may be implemented using file server manager 102. The database 306 may be used to implement and/or may be implemented using registration information 144 and/or metadata 136. The virtualized file server 334, virtualized file server 336, and virtualized file server 338 may be used to implement and/or implemented by virtualized file server 106, virtualized file server 114, and virtualized file server 122, respectively.

[0073] While three virtualized file servers are shown in FIG. 3, generally any number may be provided and may be in communication with file server manager 304. Each virtualized file server in FIG. 3 is shown as providing 2 file systems. However, generally, any number of file systems may be provided by a particular virtualization environment and/or virtualized file server—including 1, 2, 3, 4, 5, 6, 7, 8, 9, or more file systems.

[0074] The example of FIG. 3 illustrates the synchronization of metadata between file systems and/or virtualized file servers and a file server manager. The file server manager 304 is in communication with database 306. While referred to as a database, the database 306 may be implemented using any data stores and/or structures, including one or more distributed databases. The file server manager 304 may maintain, access, write to and/or read from the database 306.

[0075] Each file system provided by a virtualized file server may have associated metadata. The file system metadata is depicted in FIG. 3 as a database associated with the file system. For example, the file system 308, 'FS1', has a database of metadata. The file system 310, 'FS2', has a database of metadata. The file system 312, 'FS3', has a database of metadata. The file system 314, 'FS4', has a database of metadata. The file system 316, 'FS5', has a database of metadata. The file system 318, 'FS6', has a database of metadata. The various file system metadata may be provided by and/or in communication with the virtualization platform hosting the file system. For example, the metadata database for file system 308 and file system 310 may be hosted by virtualization platform 320—e.g., may be stored in and/or distributed across storage devices of a storage pool provided by the virtualization platform 320, which may include the storage of one or more computing nodes used to host FSVMs 326 or other components providing the virtualized file server. The metadata database for file system 312 and file system 314 may be hosted by

virtualized file server 336—e.g., may be stored in and/or distributed across storage devices of a storage pool provided by the virtualized file server 336, which may include the storage of one or more computing nodes used to host FSVMs 328. The metadata database for file system 316 and file system 318 may be hosted by virtualized file server 338—e.g., may be stored in and/or distributed across storage devices of a storage pool provided by the virtualized file server 338, which may include the storage of one or more computing nodes used to host FSVMs 330. Note that, while the file systems are shown as separate boxes in FIG. 3, the file systems may be implemented using one or more FSVMs and virtualized storage provided by the virtualization environment.

[0076] The database 306 may be used to store metadata and maintain synchronization between the metadata and the metadata of each of the file systems and/or virtualized file servers in communication with the file server manager 304. The file server manager 304 and/or admin system 302 may provide one or more synchronization processes which conduct synchronization between the database 306 and the metadata associated with file system 308, file system 310, file system 312, file system 314, file system 316, and/or file system 318. As metadata is created, destroyed, and/or changed at the virtualized file servers, the changes may be synchronized with the database 306. In this manner, the file server manager 304 may maintain access to accurate information regarding the metadata of one or more connected virtualized file servers. The metadata may be used to administer, manage, and/or access the virtualized file servers and/or file systems. Examples of metadata stored by a virtualized file server and/or file system may include, for example, authentication information for the virtualized file server and/or virtual machines in the virtualized file server, authorization information for the virtualized file server, file system, and/or virtual machines in the virtualized file server, configuration information for the virtualized file server and/or virtual machines in the virtualized file server, end point information (e.g., supported API calls and/or endpoints), a number of shares stored in the virtualized storage of the virtualized file server and/or file system, identities of the shares stored in the virtualized storage of the virtualized file server, a number of computing nodes and/or file server virtual machines (FSVMs) present in the virtualized file server, a number of files and/or directories hosted by the virtualized file server and/or file system, compute resources available and/or used at the virtualized file server, storage resources available and/or used at the virtualized file server, or other metadata regarding the virtualized file server.

[0077] Accordingly, in FIG. 3, the various virtualized file server databases containing metadata regarding their file system are shown connected to the database 306 of the file server manager 304. As changes in the metadata stored at the various virtualized file servers occur, they may be synchronized with the database 306 of the file server manager 304. The metadata may be communicated, for example over a network, between the virtualized file servers and the file server manager 304 for storage in the database 306. Synchronization may include periodic updates in some examples.

[0078] In operation, the file server manager 304 may access the database 306 to access metadata regarding one or more of the virtualized file servers to aid in managing,

accessing, and/or displaying or otherwise reporting on the status of the virtualized file servers.

[0079] FIG. 4 is a schematic illustration of a security workflow arranged in accordance with examples described herein. FIG. 4 illustrates components and operations performed by client 402, client 404, file server manager 406, and/or virtualized file servers 408. The file server manager 406 may provide API gateway 410 and authentication process 412. One or more virtualized file servers 408 may each provide an API gateway 414 and file services 416. The various operations in a security workflow are shown as numbered operations occurring between components as shown. The workflow of FIG. 4 is exemplary, and additional, fewer, and/or different components and/or operations may be used in other examples.

[0080] The file server manager 406 may be used to implement and/or may be implemented by other file server managers described herein, such as file server manager 102 of FIG. 1, and/or file server manager 304 of FIG. 3. Similarly, the virtualized file servers 408 may be used to implement and/or may be implemented by virtualized file servers described herein, such as virtualized file server 106, virtualized file server 114, and/or virtualized file server 122 of FIG. 1 or virtualized file server 334, virtualized file server 336, and/or virtualized file server 338 of FIG. 3.

[0081] During operation of a security workflow, a client may provide a request to a file server manager, such as file server manager 406. The request may be provided to a gateway, such as an API gateway of the file server manager 406, such as API gateway 410. The client may be, for example, an administrator, a process (e.g., an automation script), a virtual machine and/or a container. In some examples, the client may be hosted by a computing node used to host a portion of the virtualized file server (e.g., a computing node having an FSVM). The request may be a management request to perform an operation on or for a virtualized file server. The request may include a request to authenticate and/or authentication credentials. The client may be authenticated by communicating with an authentication process provided by the file server manager, such as authentication process 412 of FIG. 4. The authentication process may be, for example, an identity access management (IAM) process. In some examples, role-based access (RBAC) may be used. The file server manager 406 may support role-based access, for example, by using a role-based authentication process.

[0082] Once authenticated, e.g., by authentication process 412, the API gateway 410 may implement a single-sign on to a file server, such as one or more of virtualized file servers 408. Accordingly, in some examples, credentials used to authenticate to the file server manager 406 may be the same used to sign on to one or more virtualized file servers. In some examples, the credentials may not be the same, however once authenticated to the file server manager 406, the file server manager 406 may select the appropriate credentials for sign-on to a virtualized file server, and provide the credentials to the virtualized file server (e.g., to one or more FSVMs, hypervisor(s), daemon(s), and/or other components of the virtualized file server). The sign-on credentials may be provided to a gateway of the virtualized file server, such as API gateway 414 of virtualized file servers 408. Once signed on, the virtualized file server may receive and process the request, e.g., the management request. As described herein, the file server manager 406 may format the request in a

manner particular to a virtualization environment of the virtualized file server. Accordingly, the file services **416** may service the request. The file services **416** may include, for example, one or more FSVMs used to implement the virtualized file server. In some examples, the file services **416** may include one or more storage controllers or other virtualization components (e.g., hypervisor(s)).

[0083] In some examples, a client may provide a request directly to one or more virtualized file servers. For example, the client **404** is shown in FIG. 4 providing a request directly to API gateway **414** of one of virtualized file servers **408**. When the virtualized file server is being managed by the file server manager **406**, the request provided directly to one or more of the virtualized file servers **408** may be validated by the file server manager **406**, and authentication token(s) provided in response. Following authentication, the request may be serviced by virtualized file server, e.g., by file services **416**.

[0084] In this manner, file server managers may provide single sign-on capabilities for one or multiple virtualized file servers. A centralized authentication process, such as an identity access management process, may manage authentication and authorization policies for one or more virtualized file servers.

[0085] FIG. 5 is a schematic illustration of a system arranged in accordance with examples described herein. The system of FIG. 5 includes an admin system **508** in communication with virtualized file server **514**, virtualized file server **516**, and virtualized file server **518**. The virtualized file server **514** includes FSVMs **502** and is hosted on virtualization platform **520**. The virtualized file server **514** provides file system **526** and file system **528**. The virtualized file server **516** includes FSVMs **504** and is hosted on virtualization platform **522**. The virtualized file server **516** provides file system **530** and file system **532**. The virtualized file server **518** includes FSVMs **506** and is hosted on virtualization platform **524**. The virtualized file server **518** provides file system **534** and file system **536**. The admin system **508** includes file server manager **510** and policies **512**. The components shown in FIG. 5 are exemplary only. Additional, fewer, and/or different components may be used in other examples.

[0086] In some examples, the system of FIG. 5 may be used to implement and/or may be implemented by the system of FIG. 1, the system of FIG. 3 or portions thereof. For example, the file server manager **510** may be implemented using file server manager **102** and/or file server manager **304**. The virtualized file server **514**, virtualized file server **516**, and virtualized file server **518** may be used to implement and/or may be implemented by virtualized file server **334**, virtualized file server **336**, and virtualized file server **338** of FIG. 3 and/or virtualized file server **106**, virtualized file server **114**, and virtualized file server **122**, respectively.

[0087] While three virtualized file servers are shown in FIG. 5, generally any number may be provided and may be in communication with file server manager **510**. Each virtualized file server in FIG. 5 is shown as providing 2 file systems. However, generally, any number of file systems may be provided by a particular virtualization environment and/or virtualized file server—including 1, 2, 3, 4, 5, 6, 7, 8, 9, or more file systems.

[0088] The example of FIG. 5 illustrates a file server manager which may implement one or more policies on

behalf of one or more virtualized file servers. In some examples, cross-server file policies may be implemented. Policies **512** are shown accessible to file server manager **510**. The policies may be generally stored on a memory and/or storage device accessible to file server manager **510**. Generally, a policy refers to a particular plan of operations to be carried out on one or more file servers under specified conditions. Policies may be provided for upgrade, scale up, scale down, redundancy, backup, and/or tiering, as examples. In FIG. 5, a policy storage is also depicted in each file system. Each file system may have its own policies, which, in some examples, may be implemented by a file server manager described herein. Policies may in some examples be communicated from the virtualized file server(s) to file server manager **510** for implementation by file server manager **510**. The file server manager **510** may access metadata of the virtualized file servers to determine if a particular specified condition was met, and if so, may provide commands (e.g., calls) to the virtualized file server(s) to implement the policy—e.g., to scale up, scale down, backup, etc. Scaling up generally refers to adding one or more additional computing node(s) and/or FSVMs to a cluster of computing nodes used to implement the virtualized file server. Local storage devices of the added computing node may be added to the virtualized file server's storage pool, in some examples, and one or more file shares may be distributed across the storage pool, including the additional computing node. Scaling down generally refers to removing one or more computing node(s) and/or FSVMs from a cluster of computing nodes used to implement the virtualized file server. Local storage devices of the removed computing node may be removed from the virtualized file server's storage pool, and one or more file shares may be redistributed across the reduced storage pool.

[0089] In some examples, the file server managers described herein, such as file server manager **510** may implement cross-file server policies. For example, because file server manager **510** is in communication with multiple virtualized file servers—e.g., virtualized file server **514**, virtualized file server **516**, and virtualized file server **518**, the file server manager **510** may implement a policy which impacts multiple virtualized file servers. For example, a policy may designate a source virtualized file server and a destination virtualized file server. For example, the file server manager **510** may designate virtualized file server **514** as a source virtualized file server and virtualized file server **516** as a destination virtualized file server. A cross-file server policy may specify that the destination virtualized file server be used to store redundant and/or backup data of the source virtualized file server. Accordingly, the file server manager **510** may communicate with virtualized file server **514** and virtualized file server **516** to implement the cross-file server policy and back-up and/or create redundant storage on virtualized file server **516** of all or portions of virtualized file server **514**. In this manner, the virtualized file servers may not need to communicate directly with one another to implement a cross-file server policy. Instead, the file server manager **510** may centrally implement the cross-file server policy.

[0090] Moreover, file server managers described herein, such as file server manager **510** may present a view of one or more virtualized file servers—e.g., virtualized file server **514**, virtualized file server **516**, and virtualized file server **518**. The view, based on metadata synchronized to a data-

base of the file server manager 510 (e.g., as shown in database 306 of FIG. 3), may be used by one or more administrators to configure cross-file server policies. For example, an administrator may view current information about storage usage in the multiple virtualized file servers to select a source file server (e.g., a file server having less available storage than another) and a destination file server (e.g., a file server having more available storage than another, such as more available storage than the source file server). The administrator may accordingly configure a cross-file server policy based on the metadata to have the source file server back-up to the destination file server. In this manner, the file server manager 510 may provide a single pane of glass management interface for managing multiple virtualized file servers.

[0091] FIG. 6 is a schematic illustration of a system arranged in accordance with examples described herein. The system of FIG. 6 includes an admin system 602 in communication with virtualized file server 614, virtualized file server 616, and virtualized file server 618. The virtualized file server 614 includes FSVMs 608 and is hosted on virtualization platform 620. The virtualized file server 616 provides file system 626 and file system 628. The virtualized file server 616 includes FSVMs 610 and is hosted on virtualization platform 622. The virtualized file server 618 provides file system 630 and file system 632. The virtualized file server 618 includes FSVMs 612 and is hosted on virtualization platform 624. The virtualized file server 618 provides file system 634 and file system 636. The admin system 602 includes file server manager 604 and user interfaces 606. The components shown in FIG. 6 are exemplary only. Additional, fewer, and/or different components may be used in other examples.

[0092] In some examples, the system of FIG. 6 may be used to implement and/or may be implemented by the system of FIG. 1, the system of FIG. 3, the system of FIG. 5, or portions thereof. For example, the file server manager 604 may be implemented using file server manager 102 and/or file server manager 304 and/or file server manager 510. The virtualized file server 614, virtualized file server 616, virtualized file server 618 may be used to implement and/or may be implemented by virtualized file server 514, virtualized file server 516, and virtualized file server 518 of FIG. 5, and/or virtualized file server 334, virtualized file server 336, and virtualized file server 338 of FIG. 3 and/or virtualized file server 106, virtualized file server 114, and virtualized file server 122, respectively.

[0093] While three virtualized file servers are shown in FIG. 6, generally any number may be provided and may be in communication with file server manager 604. Each virtualized file server in FIG. 6 is shown as providing 2 file systems. However, generally, any number of file systems may be provided by a particular virtualization environment and/or virtualized file server—including 1, 2, 3, 4, 5, 6, 7, 8, 9, or more file systems.

[0094] The example of FIG. 6 illustrates a file server manager which may provide user interfaces to multiple virtualized file servers. The user interfaces 606 may include an interface to each of the file systems on virtualized file server 614, virtualized file server 616, and virtualized file server 618. For example, user interfaces 606 may include user interfaces to file system 626, file system 628, file system 630, file system 632, file system 634, and/or file system 636. Specifics of the file system's user interface may be provided

to the file server manager 604 during registration and/or during metadata synchronization processes.

[0095] During operation, the file server manager 604 may present a view of multiple virtualized file servers to an administrator or other user. The view may be, for example, a display, a list, or other representation of the virtualized file servers in communication with the file server manager 604. A user may select or otherwise indicate a particular virtualized file server, and responsive to the selection, the file server manager 604 may display a user interface for the selected virtualized file server. The user interface for a selected virtualized file server may include a variety of selections and options specific to the virtualized file server. An administrator or other user may use the user interface specific to the virtualized file server, and commands or other selections received through the interface may be received by the file server manager 604 and communicated to the selected virtualized file server—e.g., as management requests formatted in accordance with the virtualization platform hosting the selected virtualized file server.

[0096] In this manner, the file server manager 604 may provide a centralized UI for a user to navigate to all connected file servers and to directly login through the file server manager 604 to the virtualized file server for more specific information if desired.

[0097] Accordingly, FIG. 3 has provided a depiction of a system where a file server manager may synchronize metadata with multiple virtualized file servers. FIG. 5 has provided depiction of a system where a file server manager may implement policies for multiple virtualized file servers, including cross-server policies. FIG. 6 has provided a depiction of a system where a file server manager may implement multiple user interfaces for virtualized file servers. It is to be understood that these various features may be combined—e.g., file server managers described herein may provide metadata synchronization, virtualized file server policy implementation, and/or multiple file server user interfaces.

[0098] In some examples, file server managers described herein may additionally or instead provide centralized life cycle management. File server managers may access information about software versions running on each of multiple virtualized file servers, including software versions of multiple components. The file server managers may provide software upgrades and/or software upgrade packages to multiple virtualized file servers, and may upgrade the multiple virtualized file servers in a manner which reduces or eliminates effects of dependences between the upgrades. For example, the virtualization environment on which each virtualized file server is hosted may place constraints on the pace or selection of upgrades. The central file server manager may be used to manage upgrades at a top level and resolve platform-based dependencies.

[0099] Examples of systems and methods described herein may include a file server manager in communication with one or more virtualized file servers. Examples of virtualized file servers which may be used to implement virtualized file servers are described in, for example, U.S. Published Patent Application 2017/0235760, entitled “Virtualized file server,” published Aug. 17, 2017 on U.S. application Ser. No. 15/422,220, filed Feb. 1, 2017, both of which documents are hereby incorporated by reference in their entirety for any purpose.

[0100] FIG. 7 is a schematic illustration of a clustered virtualization environment 700 implementing a virtualized

file server (VFS 732) according to particular embodiments. In particular embodiments, the VFS 732 provides file services to user VMs 714, 718, 722, 726, 730, and 734. Each user VM may be a client as used herein. The file services may include storing and retrieving data persistently, reliably, and efficiently. The user virtual machines may execute user processes, such as office applications or the like, on host machines 702, 708, and 716. The stored data may be represented as a set of storage items, such as files organized in a hierarchical structure of folders (also known as directories), which can contain files and other folders, and shares, which can also contain files and folders.

[0101] The clustered virtualization environment 700 and/or VFS 732 may be used to implement one or more virtualization platforms and/or virtualized file servers described herein, such as the virtualized file server 106, virtualized file server 114, and/or virtualized file server 122 of FIG. 1 and/or the virtualized file server 334, virtualized file server 336, and/or virtualized file server 338 of FIG. 3 and/or any other virtualized file server described herein.

[0102] The architectures of FIG. 7 can be implemented for a distributed platform that contains multiple host machines 702, 716, and 708 that manage multiple tiers of storage. The multiple tiers of storage may include storage that is accessible through network 754, such as, by way of example and not limitation, cloud storage 706 (e.g., which may be accessible through the Internet), network-attached storage 710 (NAS) (e.g., which may be accessible through a LAN), or a storage area network (SAN). Examples described herein also permit local storage 748, 750, and 752 that is incorporated into or directly attached to the host machine and/or appliance to be managed as part of storage pool 756. Examples of such local storage include Solid State Drives (henceforth “SSDs”), Hard Disk Drives (henceforth “HDDs” or “spindle drives”), optical disk drives, external drives (e.g., a storage device connected to a host machine via a native drive interface or a serial attached SCSI interface), or any other direct-attached storage. These storage devices, both direct-attached and network-accessible, collectively form storage pool 756. Virtual disks (or “vDisks”) may be structured from the physical storage devices in storage pool 756. As used herein, the term vDisk refers to the storage abstraction that is exposed by a component of the virtualization platform, such as a Controller/Service VM (CVM) (e.g., CVM 736) and/or a hypervisor or other storage controller to be used by a user VM (e.g., user VM 714). In particular embodiments, the vDisk may be exposed via iSCSI (“internet small computer system interface”) or NFS (“network filesystem”) and is mounted as a virtual disk on the user VM. In particular embodiments, vDisks may be organized into one or more volume groups (VGs).

[0103] Each host machine 702, 716, 708 may run virtualization software, such as VMWARE ESX(I), MICROSOFT HYPER-V, or REDHAT KVM. The virtualization software includes hypervisors 742, 744, and 746 to create, manage, and destroy user VMs, as well as managing the interactions between the underlying hardware and user VMs. User VMs may run one or more applications that may operate as “clients” with respect to other elements within clustered virtualization environment 700. A hypervisor may connect to network 754. In particular embodiments, a host machine 702, 708, or 716 may be a physical hardware computing device; in particular embodiments, a host machine 702, 708, or 716 may be a virtual machine.

[0104] CVMs 736, 738, and 740 are used to manage storage and input/output (“I/O”) activities according to particular embodiments. These special VMs act as the storage controller in the currently described architecture. Multiple such storage controllers may coordinate within a cluster to form a unified storage controller system. CVMs may run as virtual machines on the various host machines, and work together to form a distributed system that manages all the storage resources, including local storage, network-attached storage 710, and cloud storage 706. The CVMs may connect to network 754 directly, or via a hypervisor. Since the CVMs run independent of hypervisors 742, 744, 746, this means that the current approach can be used and implemented within any virtual machine architecture, since the CVMs of particular embodiments can be used in conjunction with any hypervisor from any virtualization vendor. In some examples, CVMs may not be used and one or more hypervisors (e.g., hypervisors 742, 744, and/or 746) may perform the functions described with respect to the CVMs. In some examples, one or more CVMs may not be present, and the hypervisor or other component hosted on the computing nodes may provide the functions attributed to the CVM herein.

[0105] A host machine may be designated as a leader node within a cluster of host machines. For example, host machine 708 may be a leader node. A leader node may have a software component designated to perform operations of the leader. For example, CVM 738 on host machine 708 may be designated to perform such operations. A leader may be responsible for monitoring or handling requests from other host machines or software components on other host machines throughout the virtualized environment. If a leader fails, a new leader may be designated. In particular embodiments, a management module (e.g., in the form of an agent) may be running on the leader node and/or in communication with the leader node or virtual machines or containers on the leader node. For example, file server managers described herein may be in communication with the leader node in some examples.

[0106] Each CVM 736, 738, and 740 exports one or more block devices or NFS server targets that appear as disks to user VMs 714, 718, 722, 726, 730, and 734. These disks are virtual, since they are implemented by the software running inside CVMs 736, 738, and 740. Thus, to user VMs, CVMs appear to be exporting a clustered storage appliance that contains some disks. All user data (including the operating system) in the user VMs may reside on these virtual disks.

[0107] Significant performance advantages can be gained by allowing the virtualization system to access and utilize local storage 748, 750, and 752 as disclosed herein. This is because I/O performance is typically much faster when performing access to local storage as compared to performing access to network-attached storage 710 across a network 754. This faster performance for locally attached storage can be increased even further by using certain types of optimized local storage devices, such as SSDs. Further details regarding methods and mechanisms for implementing the virtualization environment illustrated in FIG. 7 are described in U.S. Pat. No. 8,601,473, which is hereby incorporated by reference in its entirety.

[0108] As a user VM performs I/O operations (e.g., a read operation or a write operation), the I/O commands of the user VM may be sent to the hypervisor that shares the same server as the user VM. For example, the hypervisor may

present to the virtual machines an emulated storage controller, receive an I/O command and facilitate the performance of the I/O command (e.g., via interfacing with storage that is the object of the command, or passing the command to a service that will perform the I/O command). An emulated storage controller may facilitate I/O operations between a user VM and a vDisk. A vDisk may present to a user VM as one or more discrete storage drives, but each vDisk may correspond to any part of one or more drives within storage pool 756. Additionally or alternatively, CVMs 736, 738, 740 may present an emulated storage controller either to the hypervisor or to user VMs to facilitate I/O operations. CVMs 736, 738, and 740 may be connected to storage within storage pool 756. CVM 736 may have the ability to perform I/O operations using local storage 748 within the same host machine 702, by connecting via network 754 to cloud storage 706 or network-attached storage 710, or by connecting via network 754 to local storage 750 or 752 within another host machine 708 or 716 (e.g., via connecting to another CVM 738 or 740). In particular embodiments, any suitable computing system may be used to implement a host machine.

[0109] In particular embodiments, the VFS 732 may include a set of File Server Virtual Machines (FSVMs) 704, 712, and 720 that execute on host machines 702, 708, and 716 and process storage item access operations requested by user VMs executing on the host machines 702, 708, and 716. The FSVMs 704, 712, and 720 may communicate with storage controllers provided by CVMs 736, 744, 740 and/or hypervisors executing on the host machines 702, 708, 716 to store and retrieve files, folders, SMB shares, or other storage items on local storage 748, 750, 752 associated with, e.g., local to, the host machines 702, 708, 716. The FSVMs 704, 712, 720 may store and retrieve block-level data on the host machines 702, 708, 716, e.g., on the local storage 748, 750, 752 of the host machines 702, 708, 716. The block-level data may include block-level representations of the storage items (e.g., files, shares). The network protocol used for communication between user VMs, FSVMs, and CVMs via the network 754 may be Internet Small Computer Systems Interface (iSCSI), Server Message Block (SMB), Network Filesystem (NFS), pNFS (Parallel NFS), or another appropriate protocol.

[0110] For the purposes of VFS 732, host machine 716 may be designated as a leader node within a cluster of host machines. In this case, FSVM 720 on host machine 716 may be designated to perform such operations. A leader may be responsible for monitoring or handling requests from FSVMs on other host machines throughout the virtualized environment. If FSVM 720 fails, a new leader may be designated for VFS 732.

[0111] In particular embodiments, the user VMs may send data to the VFS 732 (e.g., to the FSVMs) using write requests, and may receive data from it using read requests. The read and write requests, and their associated parameters, data, and results, may be sent between a user VM and one or more file server VMs (FSVMs) located on the same host machine as the user VM or on different host machines from the user VM. The read and write requests may be sent between host machines 702, 708, 716 via network 754, e.g., using a network communication protocol such as iSCSI, CIFS, SMB, TCP, IP, or the like. When a read or write request is sent between two VMs located on the same one of the host machines 702, 708, 716 (e.g., between the user VM

714 and the FSVM 704 located on the host machine 702), the request may be sent using local communication within the host machine 702 instead of via the network 754. As described above, such local communication may be substantially faster than communication via the network 754. The local communication may be performed by, e.g., writing to and reading from shared memory accessible by the user VM 714 and the FSVM 704, sending and receiving data via a local “loopback” network interface, local stream communication, or the like.

[0112] In particular embodiments, the storage items stored by the VFS 732, such as files and folders, may be distributed amongst multiple FSVMs 704, 712, 720. In particular embodiments, when storage access requests are received from the user VMs, the VFS 732 identifies FSVMs 704, 712, 720 at which requested storage items, e.g., folders, files, or portions thereof, are stored, and directs the user VMs to the locations of the storage items. The FSVMs 704, 712, 720 may maintain a storage map, such as a sharding map, that maps names or identifiers of storage items to their corresponding locations. The storage map may be a distributed data structure of which copies are maintained at each FSVM 704, 712, 720 and accessed using distributed locks or other storage item access operations. Alternatively, the storage map may be maintained by an FSVM at a leader node such as the FSVM 720, and the other FSVMs 704 and 712 may send requests to query and update the storage map to the leader FSVM 720. Other implementations of the storage map are possible using appropriate techniques to provide asynchronous data access to a shared resource by multiple readers and writers. The storage map may map names or identifiers of storage items in the form of text strings or numeric identifiers, such as folder names, file names, and/or identifiers of portions of folders or files (e.g., numeric start offset positions and counts in bytes or other units) to locations of the files, folders, or portions thereof. Locations may be represented as names of FSVMs, e.g., “FSVM-1”, as network addresses of host machines on which FSVMs are located (e.g., “ip-addr1” or 128.1.1.10), or as other types of location identifiers.

[0113] When a user application executing in a user VM 714 on one of the host machines 702 initiates a storage access operation, such as reading or writing data, the user VM 714 may send the storage access operation in a request to one of the FSVMs 704, 712, 720 on one of the host machines 702, 708, 716. A FSVM 712 executing on a host machine 708 that receives a storage access request may use the storage map to determine whether the requested file or folder is located on the FSVM 712. If the requested file or folder is located on the FSVM 712, the FSVM 712 executes the requested storage access operation. Otherwise, the FSVM 712 responds to the request with an indication that the data is not on the FSVM 712, and may redirect the requesting user VM 714 to the FSVM on which the storage map indicates the file or folder is located. The client may cache the address of the FSVM on which the file or folder is located, so that it may send subsequent requests for the file or folder directly to that FSVM.

[0114] As an example and not by way of limitation, the location of a file or a folder may be pinned to a particular FSVM 704 by sending a file service operation that creates the file or folder to a CVM 736 and/or hypervisor 742 associated with (e.g., located on the same host machine as) the FSVM 704. The CVM 736 subsequently processes file

service commands for that file for the FSVM **704** and sends corresponding storage access operations to storage devices associated with the file. The CVM **736** may associate local storage **748** with the file if there is sufficient free space on local storage **748**. Alternatively, the CVM **736** may associate a storage device located on another host machine **702**, e.g., in local storage **750**, with the file under certain conditions, e.g., if there is insufficient free space on the local storage **748**, or if storage access operations between the CVM **736** and the file are expected to be infrequent. Files and folders, or portions thereof, may also be stored on other storage devices, such as the network-attached storage (NAS) network-attached storage **710** or the cloud storage **706** of the storage pool **756**.

[0115] In particular embodiments, a name service **724**, such as that specified by the Domain Name System (DNS) Internet protocol, may communicate with the host machines **702, 708, 716** via the network **754** and may store a database of domain name (e.g., host name) to IP address mappings. The domain names may correspond to FSVMs, e.g., fsvm1.domain.com or ip-addr1.domain.com for an FSVM named FSVM-1. The name service **724** may be queried by the user VMs to determine the IP address of a particular host machine **702, 708, 716** given a name of the host machine, e.g., to determine the IP address of the host name ip-addr1 for the host machine **702**. The name service **724** may be located on a separate server computer system or on one or more of the host machines **702, 708, 716**. The names and IP addresses of the host machines of the VFS **732**, e.g., the host machines **702, 708, 716**, may be stored in the name service **724** so that the user VMs may determine the IP address of each of the host machines **702, 708, 716**, or FSVMs **704, 712, 720**. The name of each VFS instance, e.g., each file system such as FS1, FS2, or the like, may be stored in the name service **724** in association with a set of one or more names that contains the name(s) of the host machines **702, 708, 716** or FSVMs **704, 712, 720** of the VFS instance VFS **732**. The FSVMs **704, 712, 720** may be associated with the host names ip-addr1, ip-addr2, and ip-addr3, respectively. For example, the file server instance name FS1.domain.com may be associated with the host names ip-addr1, ip-addr2, and ip-addr3 in the name service **724**, so that a query of the name service **724** for the server instance name “FS1” or “FS1.domain.com” returns the names ip-addr1, ip-addr2, and ip-addr3. As another example, the file server instance name FS1.domain.com may be associated with the host names fsvm-1, fsvm-2, and fsvm-3. Further, the name service **724** may return the names in a different order for each name lookup request, e.g., using round-robin ordering, so that the sequence of names (or addresses) returned by the name service for a file server instance name is a different permutation for each query until all the permutations have been returned in response to requests, at which point the permutation cycle starts again, e.g., with the first permutation. In this way, storage access requests from user VMs may be balanced across the host machines, since the user VMs submit requests to the name service **724** for the address of the VFS instance for storage items for which the user VMs do not have a record or cache entry, as described below.

[0116] In particular embodiments, each FSVM may have two IP addresses: an external IP address and an internal IP address. The external IP addresses may be used by SMB/CIFS clients, such as user VMs, to connect to the FSVMs. The external IP addresses may be stored in the name service

724. The IP addresses ip-addr1, ip-addr2, and ip-addr3 described above are examples of external IP addresses. The internal IP addresses may be used for iSCSI communication to CVMs and/or hypervisors, e.g., between the FSVMs **704, 712, 720** and the CVMs **736, 744, 740** and/or hypervisors **742, 744**, and/or **746**. Other internal communications may be sent via the internal IP addresses as well, e.g., file server configuration information may be sent from the CVMs to the FSVMs using the internal IP addresses, and the CVMs may get file server statistics from the FSVMs via internal communication as needed.

[0117] Since the VFS **732** is provided by a distributed set of FSVMs **704, 712, 720**, the user VMs that access particular requested storage items, such as files or folders, do not necessarily know the locations of the requested storage items when the request is received. A distributed file system protocol, e.g., MICROSOFT DFS or the like, is therefore used, in which a user VM **714** may request the addresses of FSVMs **704, 712, 720** from a name service **724** (e.g., DNS). The name service **724** may send one or more network addresses of FSVMs **704, 712, 720** to the user VM **714**, in an order that changes for each subsequent request. These network addresses are not necessarily the addresses of the FSVM **712** on which the storage item requested by the user VM **714** is located, since the name service **724** does not necessarily have information about the mapping between storage items and FSVMs **704, 712, 720**. Next, the user VM **714** may send an access request to one of the network addresses provided by the name service, e.g., the address of FSVM **712**. The FSVM **712** may receive the access request and determine whether the storage item identified by the request is located on the FSVM **712**. If so, the FSVM **712** may process the request and send the results to the requesting user VM **714**. However, if the identified storage item is located on a different FSVM **720**, then the FSVM **712** may redirect the user VM **714** to the FSVM **720** on which the requested storage item is located by sending a “redirect” response referencing FSVM **720** to the user VM **714**. The user VM **714** may then send the access request to FSVM **720**, which may perform the requested operation for the identified storage item.

[0118] A particular virtualized file server, such as VFS **732**, including the items it stores, e.g., files and folders, may be referred to herein as a VFS “instance” and/or a file system and may have an associated name, e.g., FS1, as described above. Although a VFS instance may have multiple FSVMs distributed across different host machines, with different files being stored on FSVMs, the VFS instance may present a single name space to its clients such as the user VMs. The single name space may include, for example, a set of named “shares” and each share may have an associated folder hierarchy in which files are stored. Storage items such as files and folders may have associated names and metadata such as permissions, access control information, size quota limits, file types, files sizes, and so on. As another example, the name space may be a single folder hierarchy, e.g., a single root directory that contains files and other folders. User VMs may access the data stored on a distributed VFS instance via storage access operations, such as operations to list folders and files in a specified folder, create a new file or folder, open an existing file for reading or writing, and read data from or write data to a file, as well as storage item manipulation operations to rename, delete, copy, or get

details, such as metadata, of files or folders. Note that folders may also be referred to herein as “directories.”

[0119] In particular embodiments, storage items such as files and folders in a file server namespace may be accessed by clients such as user VMs by name, e.g., “\Folder-1\File-1” and “\Folder-2\File-2” for two different files named File-1 and File-2 in the folders Folder-1 and Folder-2, respectively (where Folder-1 and Folder-2 are sub-folders of the root folder). Names that identify files in the namespace using folder names and file names may be referred to as “path names.” Client systems may access the storage items stored on the VFS instance by specifying the file names or path names, e.g., the path name “\Folder-1\File-1”, in storage access operations. If the storage items are stored on a share (e.g., a shared drive), then the share name may be used to access the storage items, e.g., via the path name “\\Share-1\Folder-1\File-1” to access File-1 in folder Folder-1 on a share named Share-1.

[0120] In particular embodiments, although the VFS instance may store different folders, files, or portions thereof at different locations, e.g., on different FSVMs, the use of different FSVMs or other elements of storage pool 756 to store the folders and files may be hidden from the accessing clients. The share name is not necessarily a name of a location such as an FSVM or host machine. For example, the name Share-1 does not identify a particular FSVM on which storage items of the share are located. The share Share-1 may have portions of storage items stored on three host machines, but a user may simply access Share-1, e.g., by mapping Share-1 to a client computer, to gain access to the storage items on Share-1 as if they were located on the client computer. Names of storage items, such as file names and folder names, are similarly location-independent. Thus, although storage items, such as files and their containing folders and shares, may be stored at different locations, such as different host machines, the files may be accessed in a location-transparent manner by clients (such as the user VMs). Thus, users at client systems need not specify or know the locations of each storage item being accessed. The VFS may automatically map the file names, folder names, or full path names to the locations at which the storage items are stored. As an example and not by way of limitation, a storage item’s location may be specified by the name, address, or identity of the FSVM that provides access to the storage item on the host machine on which the storage item is located. A storage item such as a file may be divided into multiple parts that may be located on different FSVMs, in which case access requests for a particular portion of the file may be automatically mapped to the location of the portion of the file based on the portion of the file being accessed (e.g., the offset from the beginning of the file and the number of bytes being accessed).

[0121] In particular embodiments, VFS 732 determines the location, e.g., FSVM, at which to store a storage item when the storage item is created. For example, a FSVM 704 may attempt to create a file or folder using a CVM 736 on the same host machine 702 as the user VM 718 that requested creation of the file, so that the CVM 736 that controls access operations to the file folder is co-located with the user VM 718. In this way, since the user VM 718 is known to be associated with the file or folder and is thus likely to access the file again, e.g., in the near future or on behalf of the same user, access operations may use local communication or short-distance communication to improve

performance, e.g., by reducing access times or increasing access throughput. If there is a local CVM on the same host machine as the FSVM, the FSVM may identify it and use it by default. If there is no local CVM on the same host machine as the FSVM, a delay may be incurred for communication between the FSVM and a CVM on a different host machine. Further, the VFS 732 may also attempt to store the file on a storage device that is local to the CVM being used to create the file, such as local storage, so that storage access operations between the CVM and local storage may use local or short-distance communication.

[0122] In particular embodiments, if a CVM is unable to store the storage item in local storage of a host machine on which an FSVM resides, e.g., because local storage does not have sufficient available free space, then the file may be stored in local storage of a different host machine. In this case, the stored file is not physically local to the host machine, but storage access operations for the file are performed by the locally-associated CVM and FSVM, and the CVM may communicate with local storage on the remote host machine using a network file sharing protocol, e.g., iSCSI, SAMBA, or the like.

[0123] In particular embodiments, if a virtual machine, such as a user VM 714, CVM 736, or FSVM 704, moves from a host machine 702 to a destination host machine 708, e.g., because of resource availability changes, and data items such as files or folders associated with the VM are not locally accessible on the destination host machine 708, then data migration may be performed for the data items associated with the moved VM to migrate them to the new host machine 708, so that they are local to the moved VM on the new host machine 708. FSVMs may detect removal and addition of CVMs (as may occur, for example, when a CVM fails or is shut down) via the iSCSI protocol or other technique, such as heartbeat messages. As another example, a FSVM may determine that a particular file’s location is to be changed, e.g., because a disk on which the file is stored is becoming full, because changing the file’s location is likely to reduce network communication delays and therefore improve performance, or for other reasons. Upon determining that a file is to be moved, VFS 732 may change the location of the file by, for example, copying the file from its existing location(s), such as local storage 748 of a host machine 702, to its new location(s), such as local storage 750 of host machine 708 (and to or from other host machines, such as local storage 752 of host machine 716 if appropriate), and deleting the file from its existing location(s). Write operations on the file may be blocked or queued while the file is being copied, so that the copy is consistent. The VFS 732 may also redirect storage access requests for the file from an FSVM at the file’s existing location to a FSVM at the file’s new location.

[0124] In particular embodiments, VFS 732 includes at least three File Server Virtual Machines (FSVMs) 704, 712, 720 located on three respective host machines 702, 708, 716. To provide high-availability, there may be a maximum of one FSVM for a particular VFS instance VFS 732 per host machine in a cluster. If two FSVMs are detected on a single host machine, then one of the FSVMs may be moved to another host machine automatically, or the user (e.g., system administrator and/or file server manager) may be notified to move the FSVM to another host machine. The user and/or file server manager may move a FSVM to another host

machine using an administrative interface that provides commands for starting, stopping, and moving FSVMs between host machines.

[0125] In particular embodiments, two FSVMs of different VFS instances may reside on the same host machine. If the host machine fails, the FSVMs on the host machine become unavailable, at least until the host machine recovers. Thus, if there is at most one FSVM for each VFS instance on each host machine, then at most one of the FSVMs may be lost per VFS per failed host machine. As an example, if more than one FSVM for a particular VFS instance were to reside on a host machine, and the VFS instance includes three host machines and three FSVMs, then loss of one host machine would result in loss of two-thirds of the FSVMs for the VFS instance, which would be more disruptive and more difficult to recover from than loss of one-third of the FSVMs for the VFS instance.

[0126] In particular embodiments, users, such as system administrators or other users of the user VMs, may expand the cluster of FSVMs by adding additional FSVMs. Each FSVM may be associated with at least one network address, such as an IP (Internet Protocol) address of the host machine on which the FSVM resides. There may be multiple clusters, and all FSVMs of a particular VFS instance are ordinarily in the same cluster. The VFS instance may be a member of a MICROSOFT ACTIVE DIRECTORY domain, which may provide authentication and other services such as name service.

[0127] FIG. 8 illustrates data flow within a clustered virtualization environment 800 implementing a VFS instance (e.g., VFS 732) in which stored items such as files and folders used by user VMs are stored locally on the same host machines as the user VMs according to particular embodiments. As described above, one or more user VMs and a Controller/Service VM and/or hypervisor may run on each host machine along with a hypervisor. As a user VM processes I/O commands (e.g., a read or write operation), the I/O commands may be sent to the hypervisor on the same server or host machine as the user VM. For example, the hypervisor may present to the user VMs a VFS instance, receive an I/O command, and facilitate the performance of the I/O command by passing the command to a FSVM that performs the operation specified by the command. The VFS may facilitate I/O operations between a user VM and a virtualized file system. The virtualized file system may appear to the user VM as a namespace of mappable shared drives or mountable network file systems of files and directories. The namespace of the virtualized file system may be implemented using storage devices in the local storage, such as disks, onto which the shared drives or network file systems, files, and folders, or portions thereof, may be distributed as determined by the FSVMs. The VFS may thus provide features disclosed herein, such as efficient use of the disks, high availability, scalability, and others. The implementation of these features may be transparent to the user VMs. The FSVMs may present the storage capacity of the disks of the host machines as an efficient, highly-available, and scalable namespace in which the user VMs may create and access shares, files, folders, and the like.

[0128] As an example, a network share may be presented to a user VM as one or more discrete virtual disks, but each virtual disk may correspond to any part of one or more virtual or physical disks within a storage pool. Additionally or alternatively, the FSVMs may present a VFS either to the

hypervisor or to user VMs of a host machine to facilitate I/O operations. The FSVMs may access the local storage via Controller/Service VMs, other storage controllers, hypervisors, or other components of the host machine. As described herein, a CVM 736 may have the ability to perform I/O operations using local storage 748 within the same host machine 702 by connecting via the network 754 to cloud storage or NAS, or by connecting via the network 754 to 750, 752 within another host machine 708, 716 (e.g., by connecting to another CVM 738, 740).

[0129] In particular embodiments, each user VM may access one or more virtual disk images stored on one or more disks of the local storage, the cloud storage, and/or the NAS. The virtual disk images may contain data used by the user VMs, such as operating system images, application software, and user data, e.g., user home folders and user profile folders. For example, FIG. 8 illustrates three virtual machine images 810, 808, 812. The virtual machine image 810 may be a file named UserVM.vmdisk (or the like) stored on disk 802 of local storage 748 of host machine 702. The virtual machine image 810 may store the contents of the user VM 714's hard drive. The disk 802 on which the virtual machine image 810 is "local to" the user VM 714 on host machine 702 because the disk 802 is in local storage 748 of the host machine 702 on which the user VM 714 is located. Thus, the user VM 714 may use local (intra-host machine) communication to access the virtual machine image 810 more efficiently, e.g., with less latency and higher throughput, than would be the case if the virtual machine image 810 were stored on disk 804 of local storage 750 of a different host machine 708, because inter-host machine communication across the network 754 would be used in the latter case. Similarly, a virtual machine image 808, which may be a file named UserVM.vmdisk (or the like), is stored on disk 804 of local storage 750 of host machine 708, and the image 808 is local to the user VM 722 located on host machine 708. Thus, the user VM 722 may access the virtual machine image 808 more efficiently than the virtual machine 718 on host machine 702, for example. In another example, the CVM 740 may be located on the same host machine 716 as the user VM 730 that accesses a virtual machine image 812 (UserVM.vmdisk) of the user VM 730, with the virtual machine image file 812 being stored on a different host machine 708 than the user VM 730 and the CVM 740. In this example, communication between the user VM 730 and the CVM 740 may still be local, e.g., more efficient than communication between the user VM 730 and a CVM 738 on a different host machine 708, but communication between the CVM 740 and the disk 804 on which the virtual machine image 812 is stored is via the network 754, as shown by the dashed lines between CVM 740 and the network 754 and between the network 754 and local storage 750. The communication between CVM 740 and the disk 804 is not local, and thus may be less efficient than local communication such as may occur between the CVM 740 and a disk 806 in local storage 752 of host machine 716. Further, a user VM 730 on host machine 716 may access data such as the virtual disk image 812 stored on a remote (e.g., non-local) disk 804 via network communication with a CVM 738 located on the remote host machine 708. This case may occur if CVM 740 is not present on host machine 716, e.g., because CVM 740 has failed, or if the FSVM 720 has been configured to communicate with 750 on host

machine **708** via the CVM **738** on host machine **708**, e.g., to reduce computational load on host machine **716**.

[0130] In particular embodiments, since local communication is expected to be more efficient than remote communication, the FSVMs may store storage items, such as files or folders, e.g., the virtual disk images, as block-level data on local storage of the host machine on which the user VM that is expected to access the files is located. A user VM may be expected to access particular storage items if, for example, the storage items are associated with the user VM, such as by configuration information. For example, the virtual disk image **810** may be associated with the user VM **714** by configuration information of the user VM **714**. Storage items may also be associated with a user VM via the identity of a user of the user VM. For example, files and folders owned by the same user ID as the user who is logged into the user VM **714** may be associated with the user VM **714**. If the storage items expected to be accessed by a user VM **714** are not stored on the same host machine **702** as the user VM **714**, e.g., because of insufficient available storage capacity in local storage **748** of the host machine **702**, or because the storage items are expected to be accessed to a greater degree (e.g., more frequently or by more users) by a user VM **722** on a different host machine **708**, then the user VM **714** may still communicate with a local CVM **736** to access the storage items located on the remote host machine **708**, and the local CVM **736** may communicate with local storage **750** on the remote host machine **708** to access the storage items located on the remote host machine **708**. If the user VM **714** on a host machine **702** does not or cannot use a local CVM **736** to access the storage items located on the remote host machine **708**, e.g., because the local CVM **736** has crashed or the user VM **714** has been configured to use a remote CVM **738**, then communication between the user VM **714** and local storage **750** on which the storage items are stored may be via a remote CVM **738** using the network **754**, and the remote CVM **738** may access local storage **750** using local communication on host machine **708**. As another example, a user VM **714** on a host machine **702** may access storage items located on a disk **806** of local storage **752** on another host machine **716** via a CVM **738** on an intermediary host machine **708** using network communication between the host machines **702** and **708** and between the host machines **708** and **716**.

[0131] FIG. 9 illustrates an example hierarchical structure of a VFS instance (e.g., a file system) in a cluster (such as a virtualized file server) according to particular embodiments. A Cluster **902** contains two VFS instances, FS1 **904** and FS2 **906**. For example, the **902** may be used to implement and/or may be implemented by a virtualized file server described herein, such as virtualized file server **334**, virtualized file server **336**, and/or virtualized file server **338** of FIG. 3. Each VFS instance as shown in FIG. 9 may be identified by a name such as “\instance”, e.g., “\FS1” for WINDOWS file systems, or a name such as “instance”, e.g., “FS1” for UNIX-type file systems. The VFS instance FS1 **904** contains shares, including Share-1 **908** and Share-2 **910**. Shares may have names such as “Users” for a share that stores user home directories, or the like. Each share may have a path name such as WFS1\Share-1 or \FS1\Users. As an example and not by way of limitation, a share may correspond to a disk partition or a pool of file system blocks on WINDOWS and UNIX-type file systems. As another example and not by way of limitation, a share may corre-

spond to a folder or directory on a VFS instance. Shares may appear in the file system instance as folders or directories to users of user VMs. Share-1 **908** includes two folders, Folder-1 **916**, and Folder-2 **918**, and may also include one or more files (e.g., files not in folders). Each folder **916**, **918** may include one or more files **922**, **924**. Share-2 **910** includes a folder Folder-3 **912**, which includes a file File-2 **914**. Each folder has a folder name such as “Folder-1”, “Users”, or “Sam” and a path name such as “\FS1\Share-1\Folder-1” (WINDOWS) or “share-1:/fs1/Users/Sam” (UNIX). Similarly, each file has a file name such as “File-1” or “Forecast.xls” and a path name such as “\FS1\Share-1\Folder-1\File-1” or “share-1:/fs1/Users/Sam/Forecast.xls”.

[0132] FIG. 10 illustrates two example host machines **1004** and **1006**, each providing file storage services for portions of two VFS instances FS1 and FS2 according to particular embodiments. The first host machine, Host-1 **702**, includes two user VMs **1008**, **1010**, a Hypervisor **1016**, a FSVM named FileServer-VM-1 (abbreviated FSVM-1) **1020**, a Controller/Service VM named CVM-1 **1024**, and local storage **1028**. Host-1’s FileServer-VM-1 **1020** has an IP (Internet Protocol) network address of 10.1.1.1, which is an address of a network interface on Host-1 **1004**. Host-1 has a hostname ip-addr1, which may correspond to Host-1’s IP address 10.1.1.1. The second host machine, Host-2 **1006**, includes two user VMs **1012**, **1014**, a Hypervisor **1018**, a File Server VM named FileServer-VM-2 (abbreviated FSVM-2) **1022**, a Controller/Service VM named CVM-2 **1026**, and local storage **1030**. Host-2’s FileServer-VM-2 **1022** has an IP network address of 10.1.1.2, which is an address of a network interface on Host-2 **1006**.

[0133] In particular embodiments, file systems FileSystem-1A **1042** and FileSystem-2A **1040** implement the structure of files and folders for portions of the FS1 and FS2 file server instances, respectively, that are located on (e.g., served by) FileServer-VM-1 **1020** on Host-1 **1004**. Other file systems on other host machines may implement other portions of the FS1 and FS2 file server instances. The file systems **1042** and **1040** may implement the structure of at least a portion of a file server instance by translating file system operations, such as opening a file, writing data to or reading data from the file, deleting a file, and so on, to disk I/O operations such as seeking to a portion of the disk, reading or writing an index of file information, writing data to or reading data from blocks of the disk, allocating or de-allocating the blocks, and so on. The file systems **1042**, **1040** may thus store their file system data, including the structure of the folder and file hierarchy, the names of the storage items (e.g., folders and files), and the contents of the storage items on one or more storage devices, such as local storage **1028**. The particular storage device or devices on which the file system data for each file system are stored may be specified by an associated file system pool (e.g., **1048** and **1050**). For example, the storage device(s) on which data for FileSystem-1A **1042** and FileSystem-2A, **1040** are stored may be specified by respective file system pools FS1-Pool-1 **1048** and FS2-Pool-2 **1050**. The storage devices for the pool may be selected from volume groups provided by CVM-1 **1024**, such as volume group VG1 **1032** and volume group VG2 **1034**. Each volume group **1032**, **1034** may include a group of one or more available storage devices that are present in local storage **1028** associated with (e.g., by iSCSI communication) the CVM-1 **1024**. The

CVM-1 **1024** may be associated with a local storage **1028** on the same host machine **702** as the CVM-1 **1024**, or with a local storage **1030** on a different host machine **1006**. The CVM-1 **1024** may also be associated with other types of storage, such as cloud storage, networked storage or the like. Although the examples described herein include particular host machines, virtual machines, file servers, file server instances, file server pools, CVMs, volume groups, and associations there between, any number of host machines, virtual machines, file servers, file server instances, file server pools, CVMs, volume groups, and any associations there between are possible and contemplated.

[0134] In particular embodiments, the file system pool **1048** may associate any storage device in one of the volume groups **1032**, **1034** of storage devices that are available in local storage **1028** with the file system FileSystem-1A **1042**. For example, the file system pool FS1-Pool-1 **1048** may specify that a disk device named hd1 in the volume group VG1 **1032** of local storage **1028** is a storage device for FileSystem-1A **1042** for file server FS1 on FSVM-1 **1020**. A file system pool FS2-Pool-2 **1050** may specify a storage device FileSystem-2A **1050** for file server FS2 on FSVM-1 **1020**. The storage device for FileSystem-2A **1040** may be, e.g., the disk device hd1, or a different device in one of the volume groups **1032**, **1034**, such as a disk device named hd2 in volume group VG2 **1034**. Each of the file systems FileSystem-1A **1042**, FileSystem-2A **1040** may be, e.g., an instance of the NTFS file system used by the WINDOWS operating system, of the UFS Unix file system, or the like. The term “file system” may also be used herein to refer to an instance of a type of file system, e.g., a particular structure of folders and files with particular names and content.

[0135] In one example, referring to FIG. 9 and FIG. 10, an FS1 hierarchy rooted at File Server FS1 **904** may be located on FileServer-VM-1 **1020** and stored in file system instance FileSystem-1A **1042**. That is, the file system instance FileSystem-1A **1042** may store the names of the shares and storage items (such as folders and files), as well as the contents of the storage items, shown in the hierarchy at and below File Server FS1 **904**. A portion of the FS1 hierarchy shown in FIG. 9, such the portion rooted at Folder-2 **918**, may be located on FileServer-VM-2 **1022** on Host-2 **1006** instead of FileServer-VM-1 **1020**, in which case the file system instance FileSystem-1B **1044** may store the portion of the FS1 hierarchy rooted at Folder-2 **918**, including Folder-3 **912**, Folder-4 **920** and File-3 **924**. Similarly, an FS2 hierarchy rooted at File Server FS2 **906** in FIG. 9 may be located on FileServer-VM-1 **1020** and stored in file system instance FileSystem-2A **1040**. The FS2 hierarchy may be split into multiple portions (not shown), such that one portion is located on FileServer-VM-1 **1020** on Host-1 **1004**, and another portion is located on FileServer-VM-2 **1022** on Host-2 **1006** and stored in file system instance FileSystem-2B **1046**.

[0136] In particular embodiments, FileServer-VM-1 (abbreviated FSVM-1) **1020** on Host-1 **1004** is a leader for a portion of file server instance FS1 and a portion of FS2, and is a backup for another portion of FS1 and another portion of FS2. The portion of FS1 for which FileServer-VM-1 **1020** is a leader corresponds to a storage pool labeled FS1-Pool-1 **1048**. FileServer-VM-1 is also a leader for FS2-Pool-2 **1050**, and is a backup (e.g., is prepared to become a leader upon request, such as in response to a failure of another FSVM)

for FS1-Pool-3 **1052** and FS2-Pool-4 **1054** on Host-2 **1006**. In particular embodiments, FileServer-VM-2 (abbreviated FSVM-2) **1022** is a leader for a portion of file server instance FS1 and a portion of FS2, and is a backup for another portion of FS1 and another portion of FS2. The portion of FS1 for which FSVM-2 **1022** is a leader corresponds to a storage pool labeled FS1-Pool-3 **1052**. FSVM-2 **1022** is also a leader for FS2-Pool-4 **1054**, and is a backup for FS1-Pool-1 **1048** and FS2-Pool-2 **1050** on Host-1 **1004**.

[0137] In particular embodiments, the file server instances FS1, FS2 provided by the FSVMs **1020** and **1022** may be accessed by user VMs **1008**, **1010**, **1012** and **1014** via a network file system protocol such as SMB, CIFS, NFS, or the like. Each FSVM **1020** and **1022** may provide what appears to client applications on user VMs **1008**, **1010**, **1012** and **1014** to be a single file system instance, e.g., a single namespace of shares, files and folders, for each file server instance. However, the shares, files, and folders in a file server instance such as FS1 may actually be distributed across multiple FSVMs **1020** and **1022**. For example, different folders in the same file server instance may be associated with different corresponding FSVMs **1020** and **1022** and CVMs **1024** and **1026** on different host machines **1004** and **1006**.

[0138] The example file server instance FS1 **904** shown in FIG. 9 has two shares, Share-1 **908** and Share-2 **910**. Share-1 **908** may be located on FSVM-1 **1020**, CVM-1 **1024**, and local storage **1028**. Network file system protocol requests from user VMs to read or write data on file server instance FS1 **904** and any share, folder, or file in the instance may be sent to FSVM-1 **1020**. FSVM-1 **1020** (or another component, such as a hypervisor in some examples) may determine whether the requested data, e.g., the share, folder, file, or a portion thereof, referenced in the request, is located on FSVM-1, and FSVM-1 is a leader for the requested data. If not, FSVM-1 may respond to the requesting User-VM with an indication that the requested data is not covered by (e.g., is not located on or served by) FSVM-1. Otherwise, the requested data is covered by (e.g., is located on or served by) FSVM-1, so FSVM-1 may send iSCSI protocol requests to a CVM that is associated with the requested data. Note that the CVM associated with the requested data may be the CVM-1 **1024** on the same host machine **702** as the FSVM-1, or a different CVM on a different host machine **1006**, depending on the configuration of the VFS. In this example, the requested Share-1 is located on FSVM-1, so FSVM-1 processes the request. To provide for path availability, multipath I/O (MPIO) may be used for communication with the FSVM, e.g., for communication between FSVM-1 and CVM-1. The active path may be set to the CVM that is local to the FSVM (e.g., on the same host machine) by default. The active path may be set to a remote CVM instead of the local CVM, e.g., when a failover occurs.

[0139] Continuing with the data request example, the associated CVM is CVM **1024**, which may in turn access the storage device associated with the requested data as specified in the request, e.g., to write specified data to the storage device or read requested data from a specified location on the storage device. In this example, the associated storage device is in local storage **1028**, and may be an HDD or SSD. CVM-1 **1024** may access the HDD or SSD via an appropriate protocol, e.g., iSCSI, SCSI, SATA, or the like. CVM **110a** may send the results of accessing local storage **1028**, e.g., data that has been read, or the status of a data write

operation, to CVM **1024** via, e.g., SATA, which may in turn send the results to FSVM-1 **1020** via, e.g., iSCSI. FSVM-1 **1020** may then send the results to user VM via SMB through the Hypervisor **1016**.

[0140] Share-2 **910** may be located on FSVM-2 **1022**, on Host-2. Network file service protocol requests from user VMs to read or write data on Share-2 may be directed to FSVM-2 **1022** on Host-2 by other FSVMs. Alternatively, user VMs may send such requests directly to FSVM-2 **1022** on Host-2, which may process the requests using CVM-2 **1026** and local storage **1030** on Host-2 as described above for FSVM-1 **1020** on Host-1.

[0141] A file server instance such as FS1 **904** in FIG. 9 may appear as a single file system instance (e.g., a single namespace of folders and files that are accessible by their names or pathnames without regard for their physical locations), even though portions of the file system are stored on different host machines. Since each FSVM may provide a portion of a file server instance, each FSVM may have one or more “local” file systems that provide the portion of the file server instance (e.g., the portion of the namespace of files and folders) associated with the FSVM.

[0142] FIG. 11 illustrates example interactions between a client **1104** and host machines **1106** and **1108** on which different portions of a VFS instance are stored according to particular embodiments. A client **1104**, e.g., an application program executing in one of the user VMs and on the host machines of a virtualized file server described herein requests access to a folder WFS1.domain.name\Share-1\Folder-3. The request may be in response to an attempt to map \FS1.domain.name\Share-1 to a network drive in the operating system executing in the user VM followed by an attempt to access the contents of Share-1 or to access the contents of Folder-3, such as listing the files in Folder-3.

[0143] FIG. 11 shows interactions that occur between the client **1104**, FSVMs **1110** and **1112** on host machines **1106** and **1108**, and a name server **1102** when a storage item is mapped or otherwise accessed. The name server **1102** may be provided by a server computer system, such as one or more of the host machines **1106**, **1108** or a server computer system separate from the host machines **1106**, **1108**. In one example, the name server **1102** may be provided by an ACTIVE DIRECTORY service executing on one or more computer systems and accessible via the network. The interactions are shown as arrows that represent communications, e.g., messages sent via the network. Note that the client **1104** may be executing in a user VM, which may be co-located with one of the FSVMs **1110** and **1112**. In such a co-located case, the arrows between the client **1104** and the host machine on which the FSVM is located may represent communication within the host machine, and such intra-host machine communication may be performed using a mechanism different from communication over the network, e.g., shared memory or inter process communication.

[0144] In particular embodiments, when the client **1104** requests access to Folder-3, a VFS client component executing in the user VM may use a distributed file system protocol such as MICROSOFT DFS, or the like, to send the storage access request to one or more of the FSVMs of FIGS. 3-4. To access the requested file or folder, the client determines the location of the requested file or folder, e.g., the identity and/or network address of the FSVM on which the file or folder is located. The client may query a domain cache of FSVM network addresses that the client has previously

identified (e.g., looked up). If the domain cache contains the network address of an FSVM associated with the requested folder name \FS1.domain.name\Share-1\Folder-3, then the client retrieves the associated network address from the domain cache and sends the access request to the network address, starting at step **1164** as described below.

[0145] In particular embodiments, at step **1164**, the client may send a request for a list of addresses of FSVMs to a name server **1102**. The name server **1102** may be, e.g., a DNS server or other type of server, such as a MICROSOFT domain controller (not shown), that has a database of FSVM addresses. At step **1148**, the name server **1102** may send a reply that contains a list of FSVM network addresses, e.g., ip-addr1, ip-addr2, and ip-addr3, which correspond to the FSVMs in this example. At step **1166**, the client **1104** may send an access request to one of the network addresses, e.g., the first network address in the list (ip-addr1 in this example), requesting the contents of Folder-3 of Share-1. By selecting the first network address in the list, the particular FSVM to which the access request is sent may be varied, e.g., in a round-robin manner by enabling round-robin DNS (or the like) on the name server **1102**. The access request may be, e.g., an SMB connect request, an NFS open request, and/or appropriate request(s) to traverse the hierarchy of Share-1 to reach the desired folder or file, e.g., Folder-3 in this example.

[0146] At step **1168**, FileServer-VM-1 **1110** may process the request received at step **1166** by searching a mapping or lookup table, such as a sharding map **1122**, for the desired folder or file. The map **1122** maps stored objects, such as shares, folders, or files, to their corresponding locations, e.g., the names or addresses of FSVMs. The map **1122** may have the same contents on each host machine, with the contents on different host machines being synchronized using a distributed data store as described below. For example, the map **1122** may contain entries that map Share-1 and Folder-1 to the File Server FSVM-1 **1110**, and Folder-3 to the File Server FSVM-3 **1112**. An example map is shown in Table 1 below. While the example of FIG. 11 is depicted and described with respect to the FSVM processing the request, in some examples, one or more other components of a virtualized system may additionally or instead process the request (e.g., a CVM and/or a hypervisor).

Stored Object	Location
Folder-1	FSVM-1
Folder-2	FSVM-1
File-1	FSVM-1
Folder-3	FSVM-3
File-2	FSVM-3

[0147] In particular embodiments, the map **1122** or **1124** may be accessible on each of the host machines. The maps may be copies of a distributed data structure that are maintained and accessed at each FSVM using a distributed data access coordinator **1126** and **1130**. The distributed data access coordinator **1126** and **1130** may be implemented based on distributed locks or other storage item access operations. Alternatively, the distributed data access coordinator **1126** and **1130** may be implemented by maintaining a master copy of the maps **1122** and **1124** at a leader node such as the host machine **1108**, and using distributed locks to access the master copy from each FSVM **1110** and **1112**.

The distributed data access coordinator **1126** and **1130** may be implemented using distributed locking, leader election, or related features provided by a centralized coordination service for maintaining configuration information, naming, providing distributed synchronization, and/or providing group services (e.g., APACHE ZOOKEEPER or other distributed coordination software). Since the map **1122** indicates that Folder-3 is located at FSVM-3 **1112** on Host-3 **1108**, the lookup operation at step **1168** determines that Folder-3 is not located at FSVM-1 on Host-1 **1106**. Thus, at step **1162** the FSVM-1 **1110** (or other component of the virtualized system) sends a response, e.g., a “Not Covered” DFS response, to the client **1104** indicating that the requested folder is not located at FSVM-1. At step **1160**, the client **1104** sends a request to FSVM-1 for a referral to the FSVM on which Folder-3 is located. FSVM-1 uses the map **1122** to determine that Folder-3 is located at FSVM-3 on Host-3 **1108**, and at step **1158** returns a response, e.g., a “Redirect” DFS response, redirecting the client **1104** to FSVM-3. The client **1104** may then determine the network address for FSVM-3, which is ip-addr3 (e.g., a host name “ip-addr3.domain.name” or an IP address, 10.1.1.3). The client **1104** may determine the network address for FSVM-3 by searching a cache stored in memory of the client **1104**, which may contain a mapping from FSVM-3 to ip-addr3 cached in a previous operation. If the cache does not contain a network address for FSVM-3, then at step **1150** the client **1104** may send a request to the name server **1102** to resolve the name FSVM-3. The name server may respond with the resolved address, ip-addr3, at step **1152**. The client **1104** may then store the association between FSVM-3 and ip-addr3 in the client’s cache.

[0148] In particular embodiments, failure of FSVMs may be detected using the centralized coordination service. For example, using the centralized coordination service, each FSVM may create a lock on the host machine on which the FSVM is located using ephemeral nodes of the centralized coordination service (which are different from host machines but may correspond to host machines). Other FSVMs may volunteer for leadership of resources of remote FSVMs on other host machines, e.g., by requesting a lock on the other host machines. The locks requested by the other nodes are not granted unless communication to the leader host machine is lost, in which case the centralized coordination service deletes the ephemeral node and grants the lock to one of the volunteer host machines and, which becomes the new leader. For example, the volunteer host machines may be ordered by the time at which the centralized coordination service received their requests, and the lock may be granted to the first host machine on the ordered list. The first host machine on the list may thus be selected as the new leader. The FSVM on the new leader has ownership of the resources that were associated with the failed leader FSVM until the failed leader FSVM is restored, at which point the restored FSVM may reclaim the local resources of the host machine on which it is located.

[0149] At step **1154**, the client **1104** may send an access request to FSVM-3 **1112** at ip-addr3 on Host-3 **1108** requesting the contents of Folder-3 of Share-1. At step **1170**, FSVM-3 **1112** queries FSVM-3’s copy of the map **1124** using FSVM-3’s instance of the distributed data access coordinator **1130**. The map **1124** indicates that Folder-3 is located on FSVM-3, so at step **1172** FSVM-3 accesses the file system **1132** to retrieve information about Folder-3 **1144**

and its contents (e.g., a list of files in the folder, which includes File-2 **1146**) that are stored on the local storage **1120**. FSVM-3 may access local storage **1120** via CVM-3 **1116**, which provides access to local storage **1120** via a volume group **1136** that contains one or more volumes stored on one or more storage devices in local storage **1120**. At step **1156**, FSVM-3 may then send the information about Folder-3 and its contents to the client **1104**. Optionally, FSVM-3 may retrieve the contents of File-2 and send them to the client **1104**, or the client **1104** may send a subsequent request to retrieve File-2 as needed.

[0150] FIG. 12 depicts a block diagram of components of a computing system in accordance with examples described herein. It should be appreciated that FIG. 12 provides only an illustration of one implementation and does not imply any limitations with regard to the environments in which different embodiments may be implemented. Many modifications to the depicted environment may be made. The computing system may be used to implement and/or may be implemented by the file server manager **102** of FIG. 1, admin system **302** of FIG. 3, admin system **508** of FIG. 5, and/or admin system **602** of FIG. 6, for example. The components shown in FIG. 12 are exemplary only, and it is to be understood that additional, fewer, and/or different components may be used in other examples.

[0151] The computing node **1200** includes one or more communications fabric(s) **1202**, which provide communications between one or more processor(s) **1204**, memory **1206**, local storage **1208**, communications unit **1210**, and/or I/O interface(s) **1212**. The communications fabric(s) **1202** can be implemented with any architecture designed for passing data and/or control information between processors (such as microprocessors, communications and network processors, etc.), system memory, peripheral devices, and any other hardware components within a system. For example, the communications fabric(s) **1202** can be implemented with one or more buses.

[0152] The memory **1206** and the local storage **1208** may be computer-readable storage media. In the example of FIG. 12, the memory **1206** includes random access memory RAM **1214** and cache **1216**. In general, the memory **1206** can include any suitable volatile or non-volatile computer-readable storage media. In this embodiment, the local storage **1208** includes an SSD **1222** and an HDD **1224**. The memory **1206** may include executable instructions for providing a file server manager **1226**. The instructions for providing a file server manager **1226** may be used to implement and/or implemented by file server manager **102** of FIG. 1, file server manager **304** of FIG. 3, file server manager **406** of FIG. 4, file server manager **510** of FIG. 5, and/or file server manager **604** of FIG. 6.

[0153] Various computer instructions, programs, files, images, etc. may be stored in local storage **1208** and/or memory **1206** for execution by one or more of the respective processor(s) **1204** via one or more memories of memory **1206**. In some examples, local storage **1208** includes a magnetic HDD **1224**. Alternatively, or in addition to a magnetic hard disk drive, local storage **1208** can include the SSD **1222**, a semiconductor storage device, a read-only memory (ROM), an erasable programmable read-only memory (EPROM), a flash memory, or any other computer-readable storage media that is capable of storing program instructions or digital information.

[0154] The media used by local storage **1208** may also be removable. For example, a removable hard drive may be used for local storage **1208**. Other examples include optical and magnetic disks, thumb drives, and smart cards that are inserted into a drive for transfer onto another computer-readable storage medium that is also part of local storage **1208**.

[0155] Communications unit **1210**, in some examples, provides for communications with other data processing systems or devices. For example, communications unit **1210** may include one or more network interface cards. Communications unit **310** may provide communications through the use of either or both physical and wireless communications links.

[0156] I/O interface(s) **1212** may allow for input and output of data with other devices that may be connected to computing node **1200**. For example, I/O interface(s) **1212** may provide a connection to external device(s) **1218** such as a keyboard, a keypad, a touch screen, and/or some other suitable input device. External device(s) **318** can also include portable computer-readable storage media such as, for example, thumb drives, portable optical or magnetic disks, and memory cards. Software and data used to practice embodiments of the present invention can be stored on such portable computer-readable storage media and can be loaded onto and/or encoded in memory **1206** and/or local storage **1208** via I/O interface(s) **1212** in some examples. I/O interface(s) **1212** may connect to a display **1220**. Display **1220** may provide a mechanism to display data to a user and may be, for example, a computer monitor.

[0157] In various examples described herein, an admin system (e.g., admin system **508** of FIG. 5) may be utilized to implement replication of a file server. Replication and disaster recovery may facilitate share-level data replication and file-server-level disaster recovery. In the event of a planned or unplanned loss of service, write access can be restored to protected shares by failing-over to a recovery site file server. Protection policies may indicate failover details, including the primary location, recovery location, and replication schedule. Rather than having a single protection policy for an entire file server, unique policies may be configured for different shares without tracking or without regard to where, physically, the shares are stored within the file system.

[0158] Data replication may be provided from a source file server to a remote file server, which may be used for failover in case of failure of the source file server. The admin system may be used to configure replication policies by, for example, selecting shares from the source file server to be replicated to the destination file server. Replication policies may further include, in various examples, selecting a replication schedule (e.g., replication frequency) for the selected shares. The source file server may be replicated by replicating each share independently based on the replication schedule. The data replication may be based on zettabyte file system (ZFS) send and receive technology using ZFS dataset snapshots. The first replication may be a full replication using a base snapshot. Subsequent replication may transfer only a data difference between the last snapshot and a new snapshot (e.g., a delta snapshot).

[0159] Replication described herein may include maintaining the standby or destination file server such that the destination file server accepts client connections and serves data over NFS and/or SMB protocols in read only. For

example, the destination file server may be used to read data from the file system. However, before a failover or activation, the destination file server may not be used to write data to the file system. In case of a failure of the primary or source file server, or other failover trigger, failover may occur by updating settings of the destination file server to accept read and write requests, which may shorten recovery time when contrasted with a typical active-passive scenario. The replication methods described herein may provide failover and fallback workflows to orchestrate recovery of the primary file server. In some examples, additional workflows may allow the destination file server to take over the identity of the primary file server.

[0160] The replication described herein may use a task based framework. A task framework may allow for tracking progress of work done (e.g., whether individual tasks have been started, completed, or exited with an error). Tasks may be used to handle user triggered operations that take a long time and are handled asynchronously. For example, a failover workflow may be a task. The replication described herein may use a job framework in addition to the task based framework. A job may be similar to a task, where a job represents a unit of system generated action. The job framework may allow for tracking progress of system generated periodic work. For example, replication may be a job.

[0161] FIG. 13 is a schematic illustration of a system **1300** configured for replication in accordance with embodiments described herein. The system **1300** includes an admin system **1302**, including a file system manager **1304**, in communication with a primary file server **1306** and a secondary file server **1308**. The primary file server **1306** and the secondary file server **1308** may use API gateways **1310** and **1320**, respectively, to communicate with the admin system **1302**. The API gateways **1310** and **1320** may additionally allow the primary file server **1306** and the secondary file server **1308**, respectively, to communicate with an active directory server (e.g., domain server) **1328**. At the primary file server **1306**, the API gateway **1310** may further communicate with a scheduler **1312** and a replicator **1314** hosted at the primary file server **1306** to perform replication. The primary file server **1306** may further include storage **1330** utilized by, for example, the replicator **1314**. The secondary file server **1308** may similarly host a scheduler **1322** and a replicator **1324**. The secondary file server **1308** may further include storage **1332** utilized by, for example, the replicator **1324**. To perform replication, the replicator **1314** at the primary file server **1306** includes a job dispatcher **1316** to dispatch replication jobs **1318** from the primary file server **1306** to the secondary file server **1308**. The replication job **1326** at the replicator **1324** of the secondary file server **1308** completes the replication.

[0162] Note that replicators, schedulers, and job dispatchers described herein may be software components. For example, the replicator **1314**, replicator **1324**, scheduler **1312**, scheduler **1322**, and/or job dispatcher **1316** may be implemented using executable instructions for performing the replicator, scheduler, and/or job dispatcher functions. The executable instructions may be executed by one or more processors (e.g., processors on computing nodes of the primary file server **1306** and/or secondary file server **1308**) to implement the scheduler, replicator, and/or job dispatcher components.

[0163] The primary file server **1306** and the secondary file server **1308** may be used to implement and/or may be

implemented by file servers and virtualized file servers described herein, such as virtualized file servers 106, 114, 122, 334, 336, 338, 408, 514, 516, 518, 614, 616, 618, and/or 732 and file servers 212, 220, 904, 906, 1042, and/or 1040. Accordingly, the primary file server 1306 and the secondary file server 1308 may include multiple host machines hosting various software to implement the file server, such as FSVMs, virtualized storage controllers, hypervisors, and the like. For example, the primary file server 1306 and the secondary file server 1308 may each include multiple host machines with FSVMs distributed across the host machines receiving file access requests for storage items of storage pools of the primary file server 1306 and the secondary file server 1308.

[0164] The admin system 1302 may be used to implement and/or may be implemented by admin systems described herein, such as admin systems 302, 508, and/or 602. The file system manager 1304 may be used to implement and/or may be implemented by file system managers described herein, such as file system managers 102, 406, 510, and/or 604. The admin system 1302 may include an instance of a supervisory process to provide a point of communication for calls coming from or to the admin system 1302 related to the primary file server 1306 and the secondary file server 1308.

[0165] The API gateway 1310 and the API gateway 1320 may run instances of the supervisory process to handle communication between FSVMs at the respective file servers and the admin system 1302. The API gateway 1310 and the API gateways 1320 may provide a single point of communication for calls coming from the admin system 1302. For example, the supervisory processes may create tasks, communicate with various components of the file servers, and performing authorizations (e.g., through AuthN and AuthZ). In some examples, where the primary file server 1306 and the secondary file server 1308 are each hosted on a plurality of host machines, the respective API gateways 1310 and 1320 may include instances distributed across the host machines of the primary file server 1306 and the secondary file server 1308. In some examples, one host machine (e.g., a leader node) may host an instance of the API gateway configured to communicate with the admin system 1302 and other host machines may host instances of the API gateway configured to communicate with the leader node. Accordingly, FSVMs hosted on various nodes of the primary file server 1306 and/or the secondary file server 1308 may send information to, and receive information from, the admin system 1302, including the file system manager 1304.

[0166] Scheduler 1312 and scheduler 1322 may be a scheduling service running on the primary file server 1306 and the secondary file server 1308, respectively. The scheduler 1312 may be a distributed scheduling service running on multiple (e.g., each) node (e.g., host machine) of a cluster hosting the primary file server 1306 and the scheduler 1322 may be a distributed scheduling service running on multiple (e.g., each) node of a cluster hosting the secondary file server 1308. The scheduler 1312 and the scheduler 1322 may be responsible for scheduling any function calls received from the admin system 1302 (e.g., from the file system manager 1304 via the API gateway) at intervals and/or at specific times and/or dates. In some examples, individual instances of the scheduler 1312 and 1322 may schedule operations for FSVMs hosted on the same host machine as the scheduler instance. For example, a replication job including snapshotting a share managed by an

FSVM at a first host machine may be scheduled by a scheduler instance at that same host machine.

[0167] In some examples, the scheduler 1312 and 1322 may further include a remote procedure call (RPC) server responsible for serving RPC requests for managing replication policies, configuration, and policies at the primary file server 1306 and the secondary file server 1308, respectively. In some examples, the primary file server 1306 and the secondary file server 1308 may communicate with one another using such RPC servers. For example, requests to create, get, update, or delete policies (e.g., replication policies) may be managed by the schedulers 1312 and 1322. For example, where the scheduler 1312 receives a create, get, update, or delete request for a replication policy from the API gateway 1310, the scheduler 1312 may function to initiate a corresponding workflow at the primary file server 1306. When initiating such workflows, the scheduler 1312 may provide, to other components of the file server 1306, additional information such as identification of a source share, target file server (e.g., secondary file server 1308) IP address (or other identifier), and/or virtual machine details (e.g., identification of FSVMs at the source and/or destination file servers, and/or maps of components of the source file server to the destination file server).

[0168] Scheduler 1312 (e.g., the scheduler at the source file server) may add scheduled entities to local storage 1330 at the primary file server 1306. Local storage 1330 may refer to any storage in a storage pool of the primary file server 1306, which may include local storage devices of one or more nodes hosting FSVMs and forming primary file server 1306. Scheduled entities may be added for each policy and each schedule associated with a policy. A scheduled entity may include schedule details (e.g., a time and/or specific interval at which to make a function call) and the function call to be made according to the schedule details.

[0169] Replicator 1314 and replicator 1324 may include a primary replicator and node replicators for each node of the primary file server 1306 and the secondary file server 1308, respectively. A primary replicator may monitor for replication jobs created at the cluster (e.g., a cluster of computing nodes hosting the primary file server 1306 or the secondary file server 1308) and may assign the jobs or portions of the jobs to the node replicators. In some examples, the node replicators may register as job pools with the primary replicator, such that the primary replicator may assign jobs to individual nodes using a job based framework.

[0170] The replicator 1314 and the replicator 1324 may be responsible for carrying out replication workflows (e.g., jobs) scheduled for completion by the scheduler 1312 and 1322, respectively. For example, the replicators may take snapshots and send ZFS data streams to remote file servers. In the example shown in FIG. 13, the replicator 1314 may execute replication job 1318 (e.g., taking a snapshot of data at the primary file server 1306) and send data to the replicator 1324 at the secondary file server 1308. The replicator 1324 may execute a replication job 1326 to receive the data from the replicator 1314 and/or to store the received data at the desired location at the secondary file server 1308. The replicator 1314 and the replicator 1324 may further implement disaster recovery workflows, in various examples. The replicators may be implemented as a per node service (e.g., with an instance located at each node of the file server) managed by, for example, a high availability (HA) service. Accordingly, when a HA event hap-

pens, a taken-over node may start one or more replicator processes, which may be terminated when a give-back (e.g., reversion to a failed node) occurs.

[0171] Storage **1330** and storage **1332** may include different storage locations accessible at the primary file server **1306** and the secondary file server **1308**, respectively. Various information and data, such as replication policies, share information for shares of the file server, mappings between the primary file server **1306** and the secondary file server **1308**, details of jobs associated with various replication policies, dataset mappings, details of tasks associated with failover and fallback workflows, states of jobs and policies, and the like. In various examples, the admin system **1302** may communicate with the primary file server **1306** to access information at storage **1330** and/or may communicate with the secondary file server **1308** to access information at storage **1332**.

[0172] The configuration of the system **1300** is exemplary and may vary in some embodiments. For example, the admin system **1302** may communicate with additional file servers. Further, each of the admin system **1302**, the primary file server **1306**, and the secondary file server **1308** may host additional components, processes, and/or instances in various examples.

[0173] FIG. 14 is a schematic illustration of a system **1400** during replication in accordance with embodiments described herein. A first file server **1406** at a first physical site **1402** presents a first file system **1410** including at least shares **1418**, **1420**, **1422**, and **1424**. The shares **1418**, **1420**, **1422**, and **1424** of the first file system **1410** may be accessed by one or more clients **1414** requesting access from the first file server **1406**. Similarly, a second file server **1408** at a second physical site **1404** presents a second file system **1412** including at least shares **1426**, **1428**, **1430**, and **1432**. The shares **1426**, **1428**, **1430**, and **1432** of the second file system **1412** may be accessed by one or more clients **1416** requesting access from the second file server **1408**.

[0174] The first file server **1406** and the second file server **1408** may be used to implement and/or may be implemented by file servers and virtualized file servers described herein, such as virtualized file servers **106**, **114**, **122**, **334**, **336**, **338**, **408**, **514**, **516**, **518**, **614**, **616**, **618**, and/or **732** and file servers **212**, **220**, **904**, **906**, **1042**, **1040**, **1306**, and/or **1308**. Accordingly, the first file server **1406** and the second file server **1408** may include multiple host machines hosting various software to implement the file server, such as FSVMs, virtualized storage controllers, and the like. The first file server **1406** and the second file server **1408** may be referred to as a file server or files instance. As used herein, an instance may be a file server and/or a clustered instance of a file server.

[0175] The file system **1410** presented by the first file server and the file system **1412** presented by the second file server may be used to implement and/or may be implemented by file systems described herein, such as file systems **308**, **310**, **312**, **314**, **316**, **318**, **526**, **528**, **530**, **532**, **534**, **536**, **626**, **628**, **630**, **632**, **634**, **636**, **904**, **1040**, **1042**, **1044**, **1046**, and/or **1132**.

[0176] As shown, the shares **1418**, **1420**, **1422**, and **1424** of the first file system **1410** correspond to shares **1426**, **1428**, **1430**, and **1432** of the second file system **1412**. However, in various examples, the first file system **1410** may include additional shares not corresponding to shares of the second file system **1412**. Similarly, the second file system **1412** may

include additional shares not corresponding to shares of the first file system **1410**. In some examples, such additional shares may be configured for replication (e.g., backup) to other file servers.

[0177] The first file server **1406** at a first site **1402** is configured for replication to the second file server **1408** at a second site **1404**. For example, shares **1418**, **1420**, and **1422** are associated with policies making the shares **1418**, **1420**, and **1422** available for read and write access at the first file server **1406**. The associated policies replicate the shares **1418**, **1420**, and **1422** to the second file server **1408** at the second site **1404** as shares **1426**, **1428**, **1430**, and **1432** of the second file server **1408**. The shares **1418**, **1420**, and **1422** are configured to provide read only access at the second file server **1408**. The second file server **1408** at the second site **1404** is also configured for replication to the first file server **1406** at the first site **1402**. For example, share **1432** is associated with a policy making the share **1432** available for read and write access at the second file server **1408**. The associated policy replicates the share **1432** to the first filer server **1406** at the first site **1402** as share **1424** of the first file server **1406**. The share **1424** is configured for read only access at the first file server **1406**.

[0178] With respect to shares **1418**, **1420**, and **1422**, the first file server **1406** may be referred to as a source file server. With respect to shares **1426**, **1428**, and **1430**, the second file server **1408** may be referred to as a target or destination file server hosting target or destination shares **1426**, **1428**, and **1430**. As used herein, a target or destination server hosts target or destination shares, which are shares on the target or destination server where data from a source file server is copied. A target or destination file server may also be referred to as a remote file server, in various examples.

[0179] With respect to shares **1418**, **1420**, and **1422** of the first file server **1406** and shares **1426**, **1428**, and **1430** of the second file server **1408**, the first file server **1406** and the second file server **1408** form an active-passive recovery environment. As used herein, an active-passive recovery environment is one where access to a primary site (e.g., the first site **1402**) is read/write and access to a remote site (e.g., the second site **1404**) is read-only for protocol clients. For example, the share **1432** at the second file server **1408** is read/write for clients (e.g., client **1416**) accessing the share **1432** at the second file server **1408** while the corresponding share **1424** at the first file server **1406** is read-only for clients (e.g., client **1414**) accessing the share **1424** at the first file server **1406**. In an active-passive environment, the passive (e.g., backup) fileserver is active before full failover and accepts client connections and serves data over various protocols (e.g., NFS/SMB protocols) in read-only. Accordingly, access to the virtualized file server in case of a disaster at or failover from a primary (e.g., source) file server may be relatively quick.

[0180] In contrast to an active-passive environment, in an active-standby environment used in many backup solutions, the remote site has to be activated (e.g., a whole file server is constructed at the remote site) on failure of the source site in order to access data (e.g., storage items at the shares). An active-passive recovery environment results in quicker failover than an active-standby recovery environment, at least because the remote site is activated prior to failover. For example, the file system **1412** at the second file server **1408** is already established including the target shares **1426**, **1428**, and **1430** during replication of the shares **1418**, **1420**,

and **1422** at the first file server **1406**. In case of failure of the first file sever **1406**, the second file server **1408** is already activated.

[0181] The first file server **1406** and the second file server **1408** may also be referred to as being in a one-way replication relationship in opposite directions. As used herein, a one-way replication relationship in opposite directions means multiple sites are writeable for different shares or targets. For example, the first file server **1406** may be writeable for shares **1418**, **1420**, and **1422**, while the second file server **1408** may be writeable for share **1432**. The first file server **1406** and the second file server **1408** are shown in a one-to-one deployment topology. As used herein, a one-to-one deployment topology is a topology where one source replicates data to only one target. In some examples, one or more shares of the first file server **1406** and/or the second file server **1408** may, instead be deployed using a one-to-many deployment topology, wherein one source replicates data to many targets. In some examples, one or more shares of the first file server **1406** and/or the second file server **1408** may be deployed using a many-to-one deployment topology, where multiple sources replicate data to a target, or a many-to-many deployment topology, where many source sites can replicate data to many targets. In some examples, the first file server **1406** and/or the second file server **1408** may be utilized in a cascaded deployment topology, where a target site for one replication relationship can also be a source site for another.

[0182] The policies associated with the shares may be configured using a file server manager (e.g., file server manager **1304**) communicating with the first file server **1406** and the second file server **1408**. A policy may include, for example, a policy name, a list of source shares for replication from the source file server, an identity of the target or destination file server, and/or a replication schedule. In some examples, the identity of the target or destination file server may include identification of a target share at the destination file server. Policies for shares of the first file server **1406** may be stored at the first file server **1406** (e.g., at local storage) and policies for shares of the second file server **1408** may be stored at the second file server **1408** (e.g., at local storage). In some examples, an admin system used to configure the replication policies may also store details of the policies.

[0183] FIG. 15 shows a schematic illustration of a system **1500** after failover to a second site **1504** in accordance with embodiments described herein. A first file server **1506** at a first physical site **1502** presents a first file system **1510** including at least shares **1518**, **1520**, **1522**, and **1524**. A second file server **1508** at a second physical site **1504** presents a second file system **1512** including at least shares **1526**, **1528**, **1530**, and **1532**. The shares **1526**, **1528**, **1530**, and **1532** of the second file system **1512** may be accessed by one or more clients **1514**, **1516** requesting access from the second file server **1508**. A file server manager **1534** in communication with the first file server **1506** and the second file server **1508** may configure failover from the first file server **1506** to the second file server **1508** in accordance with policies associated with the shares **1518**, **1520**, **1522**, and **1524** at the first file server **1506** and/or policies associated with the shares **1526**, **1528**, **1530**, and **1532** at the second file server **1508**.

[0184] The first file server **1506** and the second file server **1508** may be used to implement and/or may be implemented by file servers and virtualized file servers described herein,

such as virtualized file servers **106**, **114**, **122**, **334**, **336**, **338**, **408**, **514**, **516**, **518**, **614**, **616**, **618**, and/or **732** and file servers **212**, **220**, **904**, **906**, **1042**, **1040**, **1306**, **1308**, **1406** and/or **1408**. Accordingly, the first file server **1506** and the second file server **1508** may include multiple host machines hosting various software to implement the file server, such as FSVMs, virtualized storage controllers, and the like.

[0185] The file system manager **1534** may be used to implement and/or may be implemented by file system managers described herein, such as file system managers **102**, **406**, **510**, **604**, and/or **1304**. The file system manager **1534** may communicate with the both the first file server **1506** and the second file server **1508** to facilitate failover from the first file server **1506** to the second file server **1508**. A user interface provided by the file system manager **1534** may be used to trigger failover from the first file server **1506** to the second file server **1508**.

[0186] The replication process described herein may support both unplanned failovers and planned failovers. Both types of failures may be triggered by the user (e.g., an admin user). An unplanned failover may, in some examples, be triggered automatically upon detection of failure of the source file server (e.g., the first file server **1506**). An unplanned failover may occur, for example, when a primary (e.g., first) site **1502** is down. The target (e.g., second) site **1504** can then be activated from the last successful snapshot per share. Clients **1514** and **1516** may then be migrated to the target site **1504** seamlessly. A planned failover may be a planned activity to change ownership of shares from the first file server **1506** to the second file server **1508**. Clients **1514** and **1516** are migrated to the second file server **1508** on failover. Planned failovers may occur without data loss when the final delta replication is done as part of the planned failover workflow.

[0187] Generally, failover from the first file server **1506** to the second file server **1508** is used to mean that the second file server **1508** is serving all client requests originally directed to the first file server **1506**. Failover can be a planned failover or an unplanned failover triggered through a communication to the file system manager **1534**. In the case of a planned failover, a request for failover may be communicated to the file system manager **1534** prior to a scheduled failover to occur at a later time. Both unplanned and planned failovers may be triggered through user interaction with a user interface presented by the file system manager **1534** and/or an admin system. Planned failovers may generally be used, for example, to conduct maintenance, upgrades, or other operations on the first file server **1506**. Unplanned failover may be used in the case of failure of the first file server **1506**.

[0188] After failover from the first file server **1506** to the second file server **1508**, the shares **1518**, **1520**, **1522**, and **1524** at the first file server **1506** may be marked as read-only shares. In some examples, the shares **1518**, **1520**, **1522**, and **1524** at the first file server **1506** may be inaccessible, and clients attempting to access the file system **1510** may be redirected to the second file server **1508**. In some examples, the shares **1518**, **1520**, **1522**, and **1524** may be configured to be inaccessible upon unplanned failover (e.g., in case of disaster at the first file server **1506**). A user requesting planned failover may, in some examples, request that the shares **1518**, **1520**, **1522**, and **1524** be configured to continue to serve read requests after failover to the second file server **1508**. The shares **1526**, **1528**, **1530**, and **1532** at the second

file server **1508** may be configured to allow for both read and write requests. For example, permissions for the shares **1526**, **1528**, **1530**, and **1532** may be changed at the file server **1508** to allow for both read and write requests by, for example, updating access information for the shares **1526**, **1528**, **1530**, and **1532**. Generally, the file system **1512** at the second file server **1508** may correspond to the file system **1510** at the first file server **1506**, meaning that the shares **1526**, **1528**, **1530**, and **1532** at the second file server **1508** correspond to the shares **1518**, **1520**, **1522**, and **1524** of the first file server **1506**.

[0189] In some examples, to complete failover from the first file server **1506** to the second file server **1508**, the file system manager **1534** may communicate with an active directory and/or domain server to update domain information such that the second file server **1508** takes over for the first file server **1506**. For example, upon failover, the file system manager **1534** may communicate with an active directory server or a domain server to assign an IP address of the first file server **1506** to the second file server **1508** such that requests directed to the IP address of the first file server **1506** are provided to the second file server **1508**. In some examples, such IP addresses may correspond to host machines and/or FSVMs of the first file server **1506** and/or the second file server **1508**. For example, consider a failover scenario where the file server **1506** had failed over to the file server **1508**. The file system manager **1534** may update a domain server and/or active directory server such that an IP address previously assigned to the file server **1506** (e.g., the IP address used to access one or more host machines of the file server **1506**) is now assigned to the file server **1508** (e.g., the IP address may be used to access one or more host machines of the file server **1508**). In this manner, when client **1514** requests access to share **1518** at the first file server **1506**, the request may be provided to the corresponding share **1526** at the second file server **1508**. Such redirection may, in various examples, be transparent or undetectable from the view of the client **1514**.

[0190] FIG. 16 illustrates an example user interface **1600** showing fileservers configured for replication (e.g., backup) and failover in accordance with examples described herein. The protected fileservers shown in the user interface **1600** may be, for example, the primary file server **1406**, the secondary file server **1408**, or other file servers in communication with a file server manager **1404**. The user interface **1600** may be displayed by the file server manager (e.g., the file server manager **102**) using a user interface **104**. The file server manager **102** may access stored information about replication policies to display the content of the user interface **1600**. For example, the file server manager **102** may maintain information about protection policies (e.g., replication policies) for shares of virtualized file servers in communication with the file server manager **102**. In some examples, the file server manager **102** may further communicate with virtualized file servers (e.g., virtualized file server **114** and/or virtualized file server **122**) to obtain additional data to display at the user interface **1600**. For example, details about replication policies may be stored at local storage of file servers in communication with the file server manager **102**.

[0191] Each row shown in the user interface **1600** may display details of replication policies applied to a particular file server in communication with the file server manager **1404**. For example, row **1602** shows that a first protected file

server is configured for replication from a source file server “LrAfsSrc” to a destination file server “LrAfsDst.” The replication is conducted in accordance with an applied policy “deptShr2P.” The user interface **1600** may also display the status of both file servers. For example, the row **1602** shows that both LrAfsSer and LrAfsDst are active (e.g., accepting read and write requests). Row **1604**, shows that a second protected file server is configured for replication from a source file server “SrcDrTest” to a destination file server “DstDrTest,” where the source file server is active and the destination file server is in standby mode (e.g., shares may be accessed on a read-only basis at the destination file server). The row **1604** shows that several policies are applied to the second source file server. For example, different policies may be used for different shares of the source file server based on desired replication frequency for the shares.

[0192] The user interface **1600** may further show actions that can be initiated with respect to each protected file server. For example, through the user interface **1600**, a user may trigger the file server manager **1404** to start, resume, and/or halt replication from the first source file server to the first destination file server. The user may also, through the user interface **1600**, trigger a failover from the second source file server to the second destination file server.

[0193] FIG. 17 illustrates an example user interface **1700** showing configuration details of a filesERVER configured for backup and failover in accordance with examples described herein. The user interface **1700** may be displayed, in some examples, using a display at an admin system (e.g., admin system **1302**) in communication with the file server. In some examples, the user interface **1700** may be displayed by selecting the name of a file server displayed by the user interface **1600**. The user interface **1700** shows details **1702** of a primary location, details **1704** of data protection policies for the relevant shares of the primary location, and details **1706** of the recovery location. For example, the details **1702** of the primary location show the name of the primary file server, the number of shares configured with a data protection or replication policy, and the active status of the primary file server. In some examples, additional details **1702** may be displayed, such as names of protected shares, physical location of the primary file server, etc. The details **1704** of the data protection policy shows the name of the policy and the configured recovery point objective for the policy. In some examples, the user interface **1700** may show details of additional data protection policies applicable to other shares of the primary file server, and/or may allow for modification of the data protection policies. Further, additional details of the replication policies may be shown. Details **1706** of the recovery file server show the target share at the recovery location, along with an indicator that the recovery file server is not in an active state. In some examples, additional details may be displayed at the user interface **1700** reflecting, for example, status of the target share, time of last update of the target share, name of the recovery file server, and the like. The information used to populate the user interface **1700** may be obtained by the file system manager from local storage of the file system manager and/or may be obtained from one or more file servers managed by the file system manager.

[0194] FIG. 18 illustrates an example user interface **1800** showing status of replication jobs in accordance with examples described herein. The user interface **1800** may be

displayed, in some examples, using a display at an admin system (e.g., admin system 1302) in communication with one or more file servers configured for share based replication in accordance with examples described herein. The table shown in the user interface 1800 may display all replication jobs for file servers configured for share based replication by the admin system 1302. The table may display, for each replication job, the start time of the replication job, the duration of the replication job, recovery point objective (RPO) compliance, name of the replicated share, name of the source file server, and/or name of the target file server. The user interface 1800 may also display a status for each replication job, reflecting that the job is, for example, in progress, has succeeded, failed, or has succeeded but is not in compliance with another criteria, such as the recovery point objective. The user interface 1800 may further include options to filter policies shown in the table by status, policy name, share, primary filter, or other criteria.

[0195] FIG. 19 illustrates an example user interface 1900 showing job details in accordance with examples described herein. The user interface 1900 may be displayed, in various examples, using a display of an admin system (e.g., admin system 1302) and may be displayed by, for example, selecting a replication job displayed in the table of replication jobs at the user interface 1800. The user interface 1900 may show further detail about a specific replication job, such as state, percentage of completion, name of the primary file server, name of the destination (e.g., recovery) file server, name of the protected share, start time of the replication job, end time of the replication job, the amount of data synced, and/or the average network throughput. Amount of data synced may refer to the quantity of data synced during the replication job. Network throughput may be displayed as an average amount of data synced per second during the replication job.

[0196] FIG. 20 illustrates an example user interface 2000 showing replication policies in accordance with examples described herein. The user interface 2000 may be displayed, in various examples, using a display of the admin system (e.g., admin system 1302). The user interface 2000 may display a table of configured policies. A new policy action element may be selectable to create a new policy. An actions menu may be selectable to display options to disable a policy, edit a policy, and/or delete a policy shown in the table of the user interface 2000. The policy information shown in the user interface 2000 may include, for example, names of shares covered by a policy, name of the policy, a primary file server hosting the shares covered by the policy, the recovery point objective of the policy, a recover file server hosting target shares for the shares covered by the policy, a number of shares covered by the policy, a status of the policy (e.g., active or inactive), and a recovery point objective compliance status. In some examples, additional information about various policies may be displayed at the user interface 2000. In various examples, the information used to populate the user interface 2000 may be obtained by the file system manager from local storage of the file system manager and/or may be obtained from one or more file servers managed by the file system manager.

[0197] FIG. 21 illustrates an example user interface 2100 showing details of a replication policy in accordance with examples described herein. The user interface 2100 may be displayed, in various examples, using a display of the admin system (e.g., admin system 1302) and may be displayed, for example, by clicking a policy name within the user interface

2100. The user interface 2100 displays a summary view for a policy. The summary view includes properties and protection status by shares tiles. The properties tile 2102 provides details including the policy description, the source file server, recovery point objective, the recovery file server, and the number of shares under the policy. The protection shares by status tile 2104 displays recovery point objective compliance for shares under the policy. The user interface 2100 may also include selectable elements allowing a user to delete or otherwise edit the replication policy. The admin system 1302 may store data (e.g., source file server identity, RPO, recovery or destination file server identity, and number of protected shares) used to populate the properties tile 2102 locally and/or may obtain such data from the source file server 1306 and/or the destination file server 1308. The admin system 1302 may further determine, based on status of replication jobs associated with the policy, whether the shares are compliant with the RPO. For example, where the RPO specifies that shares be backed up every 2 minutes and replication is occurring at the source file server 1306 at an interval longer than 2 minutes, the shares may be out of compliance with the stated RPO, as shown in the protection status by shares tile 2104.

[0198] FIG. 22 illustrates an example user interface 2200 showing replication history of a share in accordance with examples described herein. The user interface 2200 may be displayed, in various examples, using a display of the admin system (e.g., admin system 1302) and may be displayed, for example, by clicking an element providing a shares view for a policy within the user interface 2100. The shares policies view displayed in the user interface 3500 displays information per share. Details include share name, the primary (e.g., source) file server, the recovery (destination) file server, the time of the last successful snapshot, and the replication history. The replication history displays a color-coded view of succeeded, in-progress, and failed replication jobs.

[0199] FIG. 23 illustrates an example user interface 2300 showing replication job history for a share in accordance with examples described herein. The user interface 2300 may be displayed, in various examples, using a display of the admin system (e.g., admin system 1302) and may be displayed, for example, by selecting a share within the user interface 2200 to provide a share details view. The share details view shown in the user interface 2300 displays a replication job table filtered for jobs of the specified share. Jobs may be filtered by job status or the primary (source) file server. The share details view also includes a summary of the share name, the primary (source) file server, the recovery (destination) file server, recovery point objective compliance, and the last successful recovery point.

[0200] FIG. 24 illustrates an example user interface 2400 for configuring share based replication in accordance with examples described herein. The user interface 2400 may be displayed, in various examples, using a display of the admin system (e.g., the admin system 1302). The user interface 2400 may be used to configure an initial replication between a source file server (e.g., source file server 1306) and a destination file server (e.g., destination file server 1308) or to resume replication after failover between the file servers. For example, the user interface 2400 allows a user to select one of two file servers to serve as the primary file server and displays data protection policies associated with the file servers. The user interface 2400 may be used to provide

active directory and domain name information and to provide active directory credentials, in various examples.

[0201] FIG. 25 illustrates an example user interface 2500 for resuming share based replication in accordance with examples described herein. The user interface 2500 may be used to configure initial replication policies between two file servers and/or to resume replication between the two file servers. The user interface 2500 may enable a user to choose a primary file server and to create a data protection policy between the two file servers.

[0202] FIG. 26 is a flowchart depicting an example method 2600 of configuring a replication policy for a share of a file server in accordance with examples described herein. Configuration of replication policies may include creation, updating, and deletion of replication policies for protection of one or more shares of a file system. Replication policies for file servers may be configured through an admin system (e.g., admin system 1302) and a file system manager (e.g., file system manager 1304) in communication with the file servers.

[0203] At block 2602, the file system manager 1304 receives a replication policy for a share of a file server via an administrative user interface. In various examples, the user interface may include fields for entering information for a particular replication policy such as, for example, a policy name, a list of shares of the source file server to be protected by the policy, identity of the destination file server, target shares of the destination file server, and/or a replication schedule. For example, a user may access the admin system 1302 and/or the file system manager 1304 to configure the replication policies by selecting a list of shares from a source file server (e.g., the source file server 1306), a target file server (e.g., the destination file server 1308), and a replication frequency. Using such share based replication, an administrative user may select individual shares for replication instead of replicating a complete file server. Further, shares can be replicated at different frequencies such that, for example, higher priority shares may be backed-up frequently and lower priority shares may be backed-up less frequently, reducing the amount of backup operations when contrasted with a full file server backup.

[0204] The file system manager 1304 validates the replication policy at block 2604. Example validation checks may verify that the source and destination file servers are compatible (e.g., configured for replication, using the same protocols, have matching authentication types, and the like), that the source shares covered by the policy do not exist in another policy as a source share or a target share, that the source and target shares are reachable, and/or that there are no conflicting tasks (e.g., another policy being created or updated) which may conflict with the pending policy. In some examples, the admin system 1302 may provide error messages to a user (e.g., through the administrator interface) if a validation check is failed.

[0205] At block 2606, the file system manager 1304 communicates the replication policy to the source file server and/or the destination file server. In some examples, the file system manager 1304 may use registration information for the source file server and the destination file server to communicate with the source file server and the destination file server, respectively. For example, the file system manager 1304 may use access credentials for the first file server 1306 to communicate with and access the first file server 1306. The file system manager 1306 may also, in some

examples, use information about the number and location of FSVMs in the first file server 1306 to communicate with the first file server 1306. For example, where a share of the first file server 1306 is hosted at a first host machine of the first file server 1306, the file server manager 1304 may communicate with an instance of the API gateway at the first host machine to configure the policy at the first file server 1306. Other registration information, such as IP addresses of components of the file servers, share locations, and the like may also be used by the file server manager 1304 to communicate a replication policy to the source file server and/or the destination file server.

[0206] In some examples, the file system manager 1304 may create the policy in a database of the admin system 1302 with a creation status of “in progress” before communicating the replication policy to the source file server. The file system manager 1304 may then send the policy to the source file server using a REST API request communicated to the API gateway of the source file server as a task for completion by the source file server. The file system manager 1304 and/or the admin system 1302 may periodically poll the creation task until completion. The status of the task at the admin system 1302 may remain as “in progress” until completion. In the event of task failure, the status may be changed to “failed” and the admin system 1302 may roll back to previous protection policies for the source file server. When the task succeeds, the file system manager 1304 may, in some examples, receive a source-target mapping (e.g., mapping 1334) from the source file server and may update the policy in the admin database with the source-target mapping. The task may then be marked as “successful” by the admin system 1302.

[0207] Once the file system manager 1304 communicates the replication policy to the source file server, the source file server may perform one or more tasks associated with policy creation at the source file server. For example, the scheduler 1312 at the source file server 1306 may run validation checks to determine whether the source volume group count (e.g., volume groups at the source file server 1306) matches the expected target volume group count. The source file server 1306 may send, via the API gateway 1310, details of the remote file server to the admin system 1302.

[0208] The source file server 1306 may further send a request (e.g., a RPC request) to the destination file server 1308 to create shares on the destination file server 1308 to implement the policy. For example, where a new policy specifies target shares at the destination file server 1308 that do not yet exist at the destination file server 1308, the source file server 1306 may communicate a request for share creation to the destination file server 1308. In some examples, the source file server 1306 and/or the destination file server 1308 may track creation of such shares and may report on successful creation or failure to successfully create the shares to the admin system 1302. For example, the destination file server 1308 may create shares with the requested attributes and may communicate to the source file server 1306 and/or the admin system 1302 when such shares are successfully created. In some examples, the admin system 1302 may generate an alert (e.g., through the administrative user interface) regarding the status of share creation at the destination file server 1308.

[0209] The source file server 1306 may further send a request (e.g., a RPC request) to the destination file server 1308 to generate a dataset map. The source file server 1306

may use the dataset map to create a mapping (e.g., mapping 1334) between shares of the source file server 1306 and shares of the destination file server 1308. The mapping may be referred to, in some examples, as a source target map and/or a source destination map. The mapping may further include a mapping between file server virtual machines of the source file server 1306 and file server virtual machines of the destination file server 1308. In some examples, the source file server 1306 may send a REST API communication to the destination file server 1308 after creating the mapping and the destination file server 1308 may create a reverse source target map at the destination file server 1308.

[0210] The completed policy may be stored at a location accessible to the source file server 1306 (e.g., local storage 1330), a location accessible to the destination file server 1308 (e.g., local storage 1332), at local storage of the admin system 1302, and/or at a shared storage location accessible by the source file server 1306, the destination file server 1308, and/or the admin system 1302. At the source file server 1306, storing the replication policy at local storage 1330 may trigger a watch callback for schedule entities and may add policies in memory for use by the scheduler 1312 during replication. In some examples, once the policy is created and the shares are successfully created at the destination file server 1308, the policy may be visible through the user interface of the admin system 1302. For example, user interface 1700 of FIG. 17 displays policy details for a protected file server.

[0211] In various examples, once a replication policy is created for shares at the primary file server 1306 to be replicated to the secondary file server 1308, the primary file server 1306 may serve client read and write requests for the covered shares. Once the shares have been replicated to corresponding target shares at the secondary file server 1308, the secondary file server 1308 may serve client read requests for the covered shares using the corresponding target shares. The secondary file server 1308 may continue such read requests before failover from the primary file server 1306 to the secondary file server 1308.

[0212] In some examples, the method 2600 may further include updating a policy created at blocks 2602, 2604, and 2606. For example, a user may request, through a user interface presented by the admin system 1302, to update one or more replication policies is use by the file servers in communication with the admin system 1302. For example, user interface 2100, which provides a detailed policy view, also provides an option to edit or update the policy. Through such a user interface, the admin system 1302 may receive an updated policy name, list of shares to be protected under the policy, identity of the destination file server 1308, replication schedule, and/or other relevant policy information. In some examples, upon receiving a request to edit and/or update a policy, the admin system 1302 may send a request to the source file server 1306 and/or the destination file server 1308 to cancel any pending jobs associated with the established policy. The admin system 1302 may then run validation checks and communicate with the source file server 1306 and/or the destination file server 1308 to create the updated policy as described, for example, with respect to blocks 2602, 2604, and 2606 of the method 2600.

[0213] In some examples, the method 2600 may further include deleting a policy created at blocks 2602, 2604, and 2606. For example, a user may request, through a user interface presented by the admin system 1302, to update one

or more replication policies is use by the file servers in communication with the admin system 1302. For example, user interface 2100, which provides a detailed policy view, also provides an option to delete an existing policy. Upon receipt of a request to delete a policy, the admin system 1302 may connect to the source file server 1306 to request deletion of the policy at the source file server 1306. The source file server 1306 may, in some examples, mark information in the source target map associated with the policy as "not in use" to allow for the policy to be recreated in the future, if desired, without reconstructing the source target map information associated with the policy. The source file server 1306 may further delete schedules associated with the policy, deleting schedules associated with policy in memory used by the scheduler. Further, all pending jobs associated with the policy may be cancelled. Upon successful deletion of the schedules associated with the policy and cancellation of pending jobs associated with the policy, the source file server 1306 may communicate successful deletion of the policy to the admin system 1302. The admin system 1302 may then send a notification of successful deletion to the user interface.

[0214] FIG. 27 is a flowchart depicting an example method 2700 of replicating a share of a file server in accordance with examples described herein. The method 2700 may be used to replicate shares of a source file server (e.g., source file server 1306) to a destination file server (e.g., destination file server 1308) using asynchronous or recovery point objective replication. With asynchronous replication, multiple shares of the source file server 1306 are replicated to one or more destination file servers based on schedules configured for each share (e.g., the recovery point objective for each share). The shares of the source file server 1306 may be replicated according to different schedules, at different frequencies. For example, higher priority shares may be replicated to one destination file server every 2 minutes, while lower priority shares may be replicated to another destination file server every 15 minutes. Accordingly, the higher priority shares may have a 2 minute recovery point objective while the lower priority shares may have a 15 minute recovery point objective. In some examples, different shares from the source file server 1306 may be replicated to the same destination file server 1308 at different frequencies.

[0215] At block 2702, a file server accesses a replication policy for a share, where the replication policy includes a replication schedule. The replication policy may be stored locally at the source file server 1306. In some examples, the source file server 1306 may access the replication policy when triggered by the scheduler 1312 to begin replication of a share of the file server 1306.

[0216] The file server creates a snapshot of the share at block 2704. In some examples, the snapshot may be a ZFS dataset snapshot. In such examples, a first replication may be a full replication using a base snapshot. Subsequent replications may then transfer only the data difference between the previous snapshot and a new snapshot (e.g., a delta snapshot). In various examples, the replicator 1314 may create snapshots using the job framework, by scheduling a replication job 1318. The replication job may store information such as source and target file server information, share information, and dataset mapping. The replication job 1318 may, in some examples, include sub-jobs. The replication job 1318 may, for example, monitor the sub-jobs used

to replicate the share. The dataset mapping of the replication job may identify one or more FSVMs of the source file server **1306** hosting at least a portion of the share to be replicated. The replication job **1318** may create sub-jobs at each FSVM, instructing the FSVMs to snapshot the share. The FSVMs may then access the data corresponding to the share using, for example, file system mappings at the FSVM.

[0217] In some examples, snapshots may be associated with expiration time metadata to indicate when a snapshot may be deleted. For example, an hourly snapshot with a retention number or 24 may have an expiration time of the current time (e.g., time of snapshot)+24 hours. Accordingly, the snapshot is not deleted until the expiration time is passed. In some examples, if a previous replication job is running, the triggered replication job may be queued such that the current running replication job will create a new job upon completion. If there is already a queued job, the previous entry may be overwritten with the new snapshot.

[0218] At block **2706**, the file server **1306** transmits the snapshot to a destination file server **1308**. In some examples, replication may include pre-negotiation between the source file server **1306** and the destination file server **1308**, which pre-negotiation may occur through direct communication and/or using the file system manager. Such pre-negotiation may determine whether the snapshot transmitted from the source file server **1306** to the destination file server **1308** is a base snapshot or an incremental replication for use on top of an existing base snapshot. Pre-negotiation may further determine a most recent common snapshot between the source file server **1306** and the destination file server **1308**. Based on the pre-negotiation, the replication may be classified as a base replication or an incremental replication. In some examples, transmission of the snapshot may occur by streaming data (e.g., using gRPC) from the source file server **1306** to the destination file server **1308**.

[0219] In some examples, such as where a share is distributed across multiple FSVMs, multiple streams may be used to transmit a snapshot from the source file server **1306** to the destination file server **1308**. The multiple streams may be managed by sub-jobs created by the replication job **1318**. For example, a first FSVM hosting a first portion of the share may stream data to a corresponding second FSVM at the destination file server **1308**. Meanwhile, a third FSVM at the source file server **1306** may stream data to a corresponding fourth FSVM at the destination file server **1308**. In some examples, the FSVMs may further communicate with virtualized storage controllers (e.g., CVMs **1024** and/or CVMs **1026**) and/or hypervisors to locate share data for replication. In such examples, the virtualized storage controllers may access specific volume groups and/or virtualized disks including share data.

[0220] Once the replication job is complete and/or all sub-jobs created by the replication job are complete, the source file server **1306** may communicate with the admin system **1302** with an updated status for the replication job. The admin system may, in some examples, display the results of the replication job, such as in user interfaces **1800** and **1900**.

[0221] FIG. 28 is a flowchart depicting an example method **2800** for unplanned failover in accordance with examples described herein. When a disaster happens on a primary file server (e.g., source file server **1306**), a user may trigger an unplanned failover for that source file server **1306**. In some examples, an unplanned failover may occur auto-

matically responsive to detected failure of the source file server **1306**. The unplanned failover may make shares on a corresponding standby file server (e.g., destination file server **1308**) available for read and write requests and may configure active directory and domain information for clients to seamlessly connect to the destination file server **1308**.

[0222] At block **2802**, the file system manager **1304** receives a request for an unplanned failover from a first file server to a second file server. The request may be received through a user interface displayed by the admin system **1302**. For example, a user may select the “failover” element of user interface **1600** to begin an unplanned failover. Other user interfaces presented by the admin system **1302** may provide alternative methods of triggering an unplanned failover. In some examples, the user may select the failed primary file server from a list of available primary file servers. Choosing a primary file server may select shares that are mapped to the standby file server (e.g., destination file server **1308**).

[0223] The file system manager **1304** updates permissions for shares at a second file server at block **2804**. Permissions may be, for example, adding permission to write to storage items at the second file server. Once a request for unplanned failover is received, the admin system **1302** may send a request to the destination file server **1308** through the API gateway **1320** to update permissions on the shares to fail over. At the destination file server **1308**, the API gateway **1320** may communicate with instances of the replicator **1324** at various nodes of the destination file server **1308** to update permissions to the shares to make the shares available for both read and write access. As part of updating permissions, FSVMs and/or the replicator **1324** may identify the last successful snapshot of a share for access. For example, if a replication job is incomplete, the replicator **1324** may roll back the share to the last complete snapshot.

[0224] The file system manager **1304** redirects client connections to the second file server at block **2806**. Clients may be redirected to the destination file server **1308** by updating domain name service mappings such that the destination file server **1308** takes over for the source file server **1306**. Redirection may further include updating service principle names of the source file server **1306** to the destination file server **1308** account in the active directory. For example, IP addresses associated with the source file server **1306** may be updated to redirect to the destination file server **1308** such that client requests for the source file server **1306** are redirected to the destination file server **1308**. In some examples, the file system manager **1304** may, in a failover request to the destination file server **1308**, request that the destination file server **1308** (e.g., the replicator **1324** at the destination file server **1308**) update such information by communicating directly with the active directory/domain server **1328**.

[0225] After the unplanned failover, the backup file server (e.g., the destination file server **1308**) is activated for all the shares, meaning that the shares are available for read and write access at the destination file server **1308**.

[0226] FIG. 29 is a flowchart depicting an example method **2900** for planned failover in accordance with examples described herein. In a planned failover, a primary file server (e.g., source file server **1306**) and a secondary file server (e.g., destination file server **1308**) may be in an active-standby configuration. That is, the source file server **1306** may serve both read and write requests for a set of

shares and the destination file server **1308** may serve read requests for a corresponding set of shares at the destination file server **1308**. A user may trigger a planned failover (e.g., through a user interface of the admin system **1302**) and choose the source file server **1306** for failover. The failover may then occur for each share of the source file server **1306** to the destination file server **1308** for the shares, as noted, for example, in a source target map at the source file server **1306**. Accordingly, in some examples, failover of a source file server **1306** may involve more than one destination file server, where some shares of the source file server **1306** are configured for replication to one destination file server and other shares of the source file server **1306** are configured for replication to another destination file server.

[0227] At block **2902**, the file system manager **1304** receives a request for planned failover from a first file server to a second file server. The request may be received through a user interface displayed by the admin system **1302**. For example, a user may select the “failover” element of user interface **1600** to begin a planned failover. Other user interfaces presented by the admin system **1302** may provide alternative methods of triggering a planned failover. In some examples, the admin system **1302** may provide a user interface to schedule a planned failover for some future time. In some examples, after receiving a request for planned failover, the file system manager **1304** may cancel future replication jobs at the destination file server **1308** and/or may notify the user of replication jobs still in progress, such that the planned failover is less likely to interrupt in-progress replication jobs.

[0228] The file system manager **1304** updates policies and permissions for shares at the first file server at block **2904**. Updating permissions may include, for example, making storage items for shares at the first file server **1306** read-only. The file system manager may update the policies and permissions for shares at the first file server by communicating with the first file server **1306** (e.g., through the API gateway **1310**). At the source file server **1306**, the API gateway **1310** may communicate with instances of the replicator **1314** at various nodes of the source file server **1306** to update permissions to the shares to make the shares available for only read access and/or to make the shares unavailable for any access, depending on user preferences and/or system configuration.

[0229] At block **2906**, the file system manager **1304** updates permissions for shares at the second file server. The admin system **1302** may send a request to the destination file server **1308** through the API gateway **1320** to update permissions on the shares to fail over. At the destination file server **1308**, the API gateway **1320** may communicate with instances of the replicator **1324** at various nodes of the destination file server **1308** to update permissions to the shares to make the shares available for both read and write access.

[0230] The file system manager **1304** redirects client connections to the second file server at block **2908**. Clients may be redirected to the destination file server **1308** by updating domain name service mappings such that the destination file server **1308** takes over for the source file server **1306**. For example, IP addresses associated with the source file server **1306** may be updated to redirect to the destination file server **1308** such that client requests for the source file server **1306** are redirected to the destination file server **1308**. In some examples, the file system manager **1304** may, in a failover

request to the destination file server **1308**, request that the destination file server **1308** (e.g., the replicator **1324** at the destination file server **1308**) update such information by communicating directly with the active directory/domain server **1328**.

[0231] FIG. **30** is a flowchart depicting an example method **3000** for resuming replication in accordance with examples described herein. A resume replication workflow may be used when a source file server (e.g., source file server **1306**) is available again after an unplanned failover. Users may have an option to choose between resuming replication from the source file server **1302** or from a destination file server (e.g., destination file server **1308**). In some examples, when the source file server **1306** comes back online after an unplanned failure or downtime, shares at the source file server **1306** and corresponding shares at the destination file server **1308** may be accessible for both read and write requests.

[0232] At block **3002**, the file system manager **1304** receives a request for resuming replication from a first file server **1306**. The request to resume replication may be received through a user interface displayed by the admin system **1302**, such as user interface **2400** and/or user interface **2500**. Other user interfaces presented by the admin system **1302** may provide alternative methods of resuming replication.

[0233] The file system manager **1304** configures replication policies at the first file server at block **3004**. Configuring replication policies may include, for example, re-activating previously deactivated replication policies at the source file server **1306**. In some examples, before configuring replication policies, the file system manager **1304** may reverse replication from the destination file server **1308** to the source file server **1306** to capture changes made to the storage items made during failover to the destination file server **1308**. The file system manager **1304** may then continue to configure replication from the source file server **1306** to the destination file server **1308**. The file system manager **1304** may configure the replication policies using one or more operations of the method **2600**.

[0234] At block **3006**, the file system manager **1304** updates policies and permissions for shares at a second file server. The file system manager **1304** may send a request to the destination file server **1308** through the API gateway **1320** to update permissions on the shares to fail over. At the destination file server **1308**, the API gateway **1320** may communicate with instances of the replicator **1324** at various nodes of the destination file server **1308** to update permissions to the shares to make the shares available for only read access.

[0235] The file system manager **1304** redirects client connections to the first file server at block **3008**. Clients may be redirected to the source file server **1306** by updating domain name service mappings such that the source file server **1306** takes over for the destination file server **1308**.

[0236] While the method **3000** is described with respect to two file servers, in various examples, failback may include additional file servers. For example, a source file server may replicate some shares to a first destination file server and some shares to a second destination file server. The source file server may, accordingly, fail over to both the first destination file server and the second destination file server. Accordingly, failback may include failback from the first destination file server and the second destination file server.

[0237] FIG. 31 is a flowchart depicting an example method 3100 of planned failback at a first file server in accordance with examples described herein. Planned failback may be a controlled operation to move ownership from a second file server (e.g., destination file server 1308) to a first file server (e.g., source file server 1306), where clients are migrated back to the first file server.

[0238] The file system manager 1304 receives a request for planned failback from a second file server to a first file server at block 3102. The request for planned failback may be received through a user interface displayed by the admin system 1302. A planned failback may be similar to a planned failover. For example, the file system manager 1304 may update permissions for the share at each of the file servers and redirect clients to the first file server from the second file server.

[0239] At block 3104, the file system manager 1304 updates policies and permissions for shares at the second file server. The file system manager 1304 may send a request to the destination file server 1308 through the API gateway 1320 to update permissions on the shares to fail over. At the destination file server 1308, the API gateway 1320 may communicate with instances of the replicator 1324 at various nodes of the destination file server 1308 to update permissions to the shares to make the shares available for only read access.

[0240] The file system manager 1304 updates permissions for shares at the first file server at block 3106. The file system manager 1304 may send a request to the source file server 1306 through the API gateway 1310 to update permissions on the shares to fail over. At the source file server 1306, the API gateway 1310 may communicate with instances of the replicator 1314 at various nodes of the source file server 1306 to update permissions to the shares to make the shares available for read and write access.

[0241] At block 3108, the file system manager 1304 redirects client connections to the first file server. Clients may be redirected to the source file server 1306 by updating domain name service mappings such that the source file server 1306 takes over for the destination file server 1308.

[0242] The above methods provide for share based replication and quick failover using active-passive backup configurations. Replication and disaster recovery may facilitate share-level data replication and file-server-level disaster recover. In the event of a planned or unplanned loss of service, write access can be restored to protected shares by failing-over to a recovery site file server. Protection policies may indicate failover details, including the primary location,

recovery location, and replication schedule. Rather than having a single protection policy for an entire file server, unique policies may be configured for different shares without tracking where, physically, the shares are stored within the file system.

[0243] From the foregoing it will be appreciated that, although specific embodiments have been described herein for purposes of illustration, various modifications may be made while remaining with the scope of the claimed technology.

[0244] Examples described herein may refer to various components as “coupled” or signals as being “provided to” or “received from” certain components. It is to be understood that in some examples the components are directly coupled one to another, while in other examples the components are coupled with intervening components disposed between them. Similarly, signal may be provided directly to and/or received directly from the recited components without intervening components, but also may be provided to and/or received from the certain components through intervening components.

1. At least one computer readable medium encoded with instructions which, when executed, cause a system to perform operations comprising:

- updating a selected share of a destination distributed file server based on a snapshot of at least a portion of a selected share of a source distributed file server, the selected share of the destination distributed file server corresponding to the selected share of the source distributed file server, wherein the selected share of the destination distributed file server is updated while the source distributed file server serves client requests for storage items of the selected share of the source distributed file server;

- receiving a request to failover from the source distributed file server to the destination distributed file server;
- configuring the destination distributed file server to service read and write requests for storage items of the selected share of the destination distributed file server, wherein the configuring is responsive to the request to failover from the source distributed file server to the destination distributed file server; and

- redirecting client requests for storage items of the selected share of the source distributed file server to the destination distributed file server by updating active directory information for the source distributed file server to reference the destination distributed file server.

* * * * *