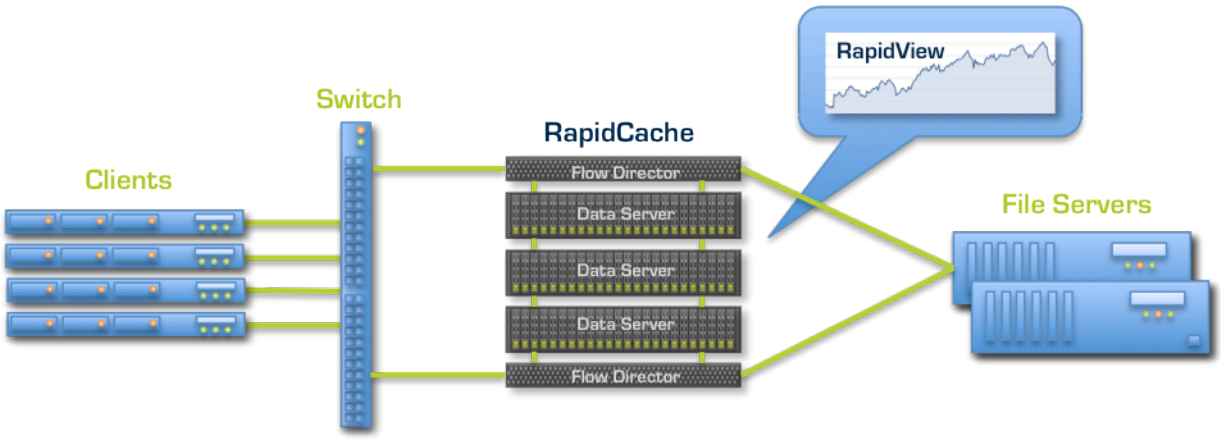
RapidCache Technical Overview

Cache IQ's RapidCache improves NFS storage performance with an in-line appliance. Without client or file server configuration changes, RapidCache intelligently finds the active dataset (or a designated dataset) and places it in DRAM and SSD. This whitepaper explains in detail how automatic storage tiering is accomplished. Furthermore, it offers specific examples of where RapidCache moves beyond NFS performance tuning and offers unprecedented control over storage resources.

RapidCache differs fundamentally from other caching solutions because it operates independently, is self-contained, and installs in the network path between the clients and file servers. This is accomplished through Cache IQ's powerful packet-inspection engine that makes real-time decisions and provides real-time visibility as it dissects every packet. Knowing the contents of each packet allows NFS data to be prioritized optimally the first time, instead of being moved after-the-fact. Packet-inspection is the foundation of RapidCache's architecture. It enables RapidCache to be dynamic, efficient, and incredibly powerful.



RapidCache Components

The RapidCache solution includes three distinct components: Data Server, Flow Director, and IQ OS. The workhorse of the RapidCache solution is the Data Server, where packet-inspection and NFS caching occur. Multiple Data Servers can be clustered and are connected to the network through the Flow Director, which interconnects the cluster and provides client and file server network connectivity.

Statistics gathered through Data Servers and Flow Directors are presented in Cache IQ's web-based analytics and configuration product, RapidView.

Data Server

The Data Server provides the packet inspection and NFS caching infrastructure for RapidCache. It is based on leading-edge commodity hardware and incorporates multiple cache media providing different performance levels. As hardware evolves, more advanced components can easily be incorporated without disruption. Currently the Data Server supports DDR3 DRAM and high performance SSD storage for caching. It communicates with both clients and file servers, inspecting every message and providing the information necessary to intelligently cache application data. The architecture is designed to integrate future Data Server models into existing clusters, taking advantage of newer hardware technology while extending the service life of previous ones.

Flow Director

The Flow Director is, at its core, an enterprise-level 10Gb Ethernet switch with custom software. It sits invisibly between clients and file servers, and has four roles within the RapidCache solution. First, it interconnects and load balances clustered nodes. Second, it is the ingress and egress point to the network. Third, it filters traffic and passes through non-supported protocol packets. And lastly, working in concert with IQ OS, the Flow Director provides the failover functionality that ensures that access to data is never interrupted.

The Flow Director, by design, does not participate in switching protocols between client and file reciprocal ports. Protocols like Spanning Tree (STP) or VLAN Trunking Protocol (VTP) pass through without interference. Furthermore, the Flow Director works with the Data Servers to support the use of Link Aggregation (LAG), 802.1Q VLAN tagging, and jumbo frames.

Supported LAG Protocols

Static LAG

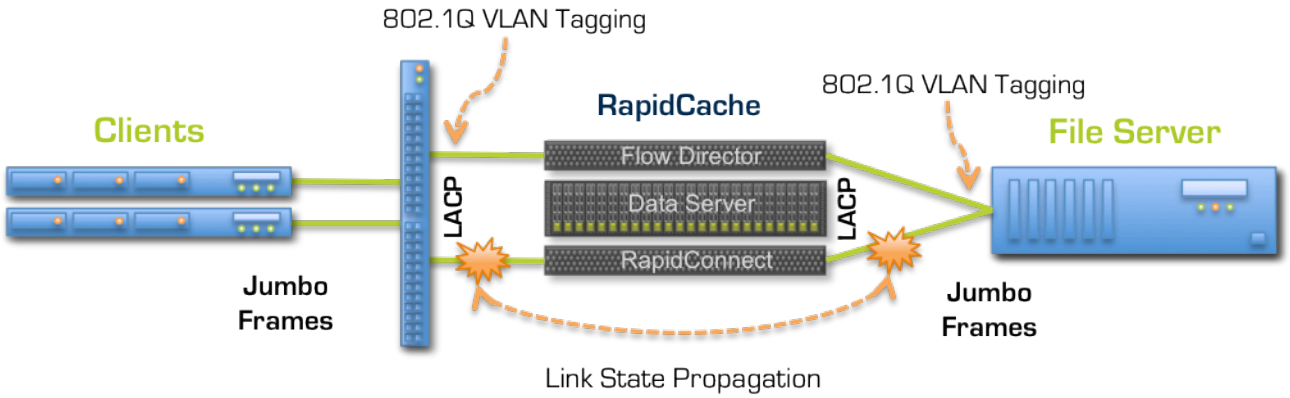
- Netapp Static Multimode
- Cisco Etherchannel: channel-group mode on

Dynamic LAG (LACP)

- Netapp Dynamic Multimode
- Cisco Etherchannel: channel-group mode active & channel-protocol LACP

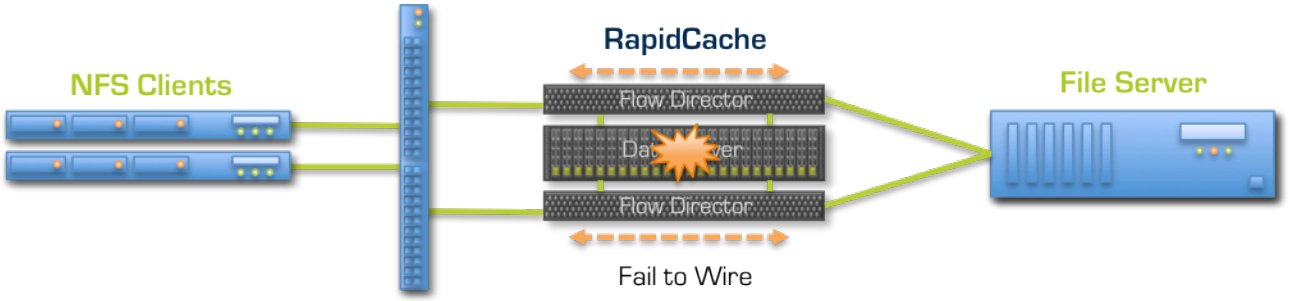
Link State Propagation

Link State Propagation (LSP) maintains end-to-end link state between the client switch and file server. Since the Flow Director sits physically between these devices, it must actively monitor reciprocal connections so both client-side and file server-side connections are in sync. This allows LAGs (if employed) to dynamically adjust in case of link failure.



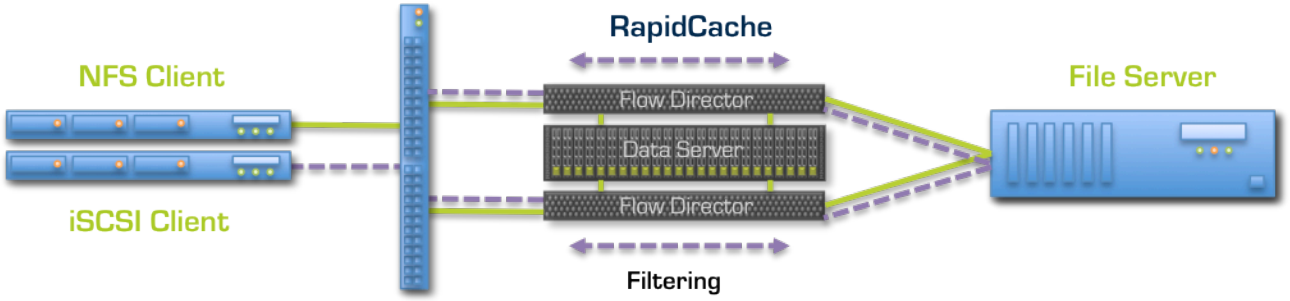
Fail to Wire

One of the most important functions of the Flow Director is the fail to wire feature that automatically bypasses the Data Servers in case of cluster failure. Using active heartbeats, the Flow Director is aware of node availability and can instantly redirect client requests to the file server when trouble is detected. Bypass mode can also be activated manually through RapidView to perform maintenance on Data Server nodes without downtime. When the administrator is ready to reactivate the cluster, cached data can be revalidated with Multi-Path Support (MPS) or flushed to start with a clear cache.



Filtering

Advanced, low-latency, wire-speed filtering occurs on the Flow Director. It passes only supported-protocol traffic to the cluster, ensuring that no latency is added to non-cached traffic and allowing the Data Servers to focus only on traffic it accelerates.



IQ OS

The IQ OS is the operating system for RapidCache. It leverages the runtime and reliability of FreeBSD, but is highly customized with a purpose built Cache IQ kernel. IQ OS operates across all Data Servers and Flow Directors and serves three basic functions: network proxy, file object server, and generic storage access. As a network proxy between clients and file servers, IQ OS performs Layer2 (L2) topology discovery to establish what is physically connected. Once the topology is understood, it maintains the network state of all connections. As requests are intercepted, they are converted to NAS-vendor independent file operations, streamlining the process while allowing RapidCache to incorporate other network protocols in the future.

Once converted, the file object server handles generic metadata operations and data operations are mapped to virtual devices. Virtual devices can be DRAM, flash memory, and other media; and are categorized according to their performance metrics, including latency and bandwidth. Virtualization of devices allows IQ OS to easily incorporate faster media to further improve performance or denser media to add cache capacity. Once the media hierarchy or tier is established, blocks are promoted and demoted based on frequency of use, unless “pinned” to a specific tier by the administrator. Additionally, the policy engine, through user-defined polices, can proactively monitor the tiers of cache and prioritize the eviction of data blocks.

Currently Available Virtual Tiers

DRAM

DRAM is currently the fastest tier of virtual media within RapidCache and is where metadata is stored for quickest random I/O. Optionally, user-defined profiles can “pin” data here for guaranteed, consistent access to critical data. SWAP files, database files, and I/O intensive virtual machine files (VMDKs) are a few examples of when pinning data in DRAM can provide unrivaled performance.

Solid State Disks (SSD)

RapidCache supports enterprise level SSDs, which can be hot added at any time to expand cache capacity. To maximize performance and capacity, individual SSDs are treated as an independent virtual tier and do not employ RAID. In the event of a failed SSD, the overall cache size will shrink by the missing SSD and will not affect others. The previously cached data will be retrieved from the file server (as requested) and stored on available media.

Future considerations for Virtual Tiers

RapidCache has the flexibility to support almost any virtual media tier including:

- PCIe Flash modules — higher throughput and lower latency
- iSCSI — higher density and easy expansion over Ethernet

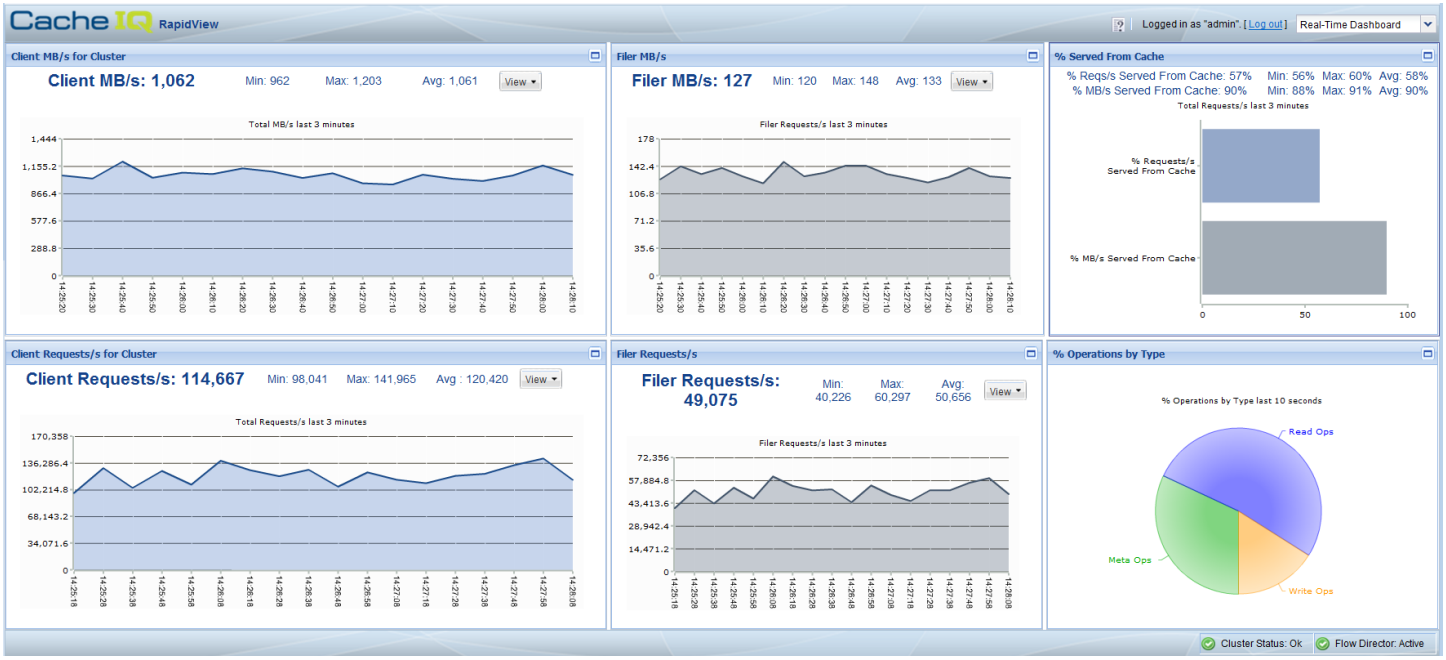
- FCOE — higher density and easy expansion over Ethernet
- New technology as it becomes available or as market demands

RapidView

The RapidView software provides a web-based user interface to view statistics and configure the cluster. It is not in the data path and is not required for normal operation. Statistics are updated in near real-time intervals of 10 seconds to aid in NFS troubleshooting and can be searched based on file server, client, filename, and time/date. After reviewing file and client trends, cache behavior can be modified by creating policies to optimize NFS performance. Additionally, custom reports can be generated for review or archive.

RapidView comes installed on a Data Server when configured as a single node or on an external server when at least two Data Server nodes are clustered. When installed on a Data Server, it is compartmentalized and uses dedicated processors and memory to prevent competition with normal Data Server functions. When clustered, an external RapidView server must be directly attached to a Flow Director to leverage the private low-latency 10Gb Ethernet network between nodes.

RapidView Screenshot

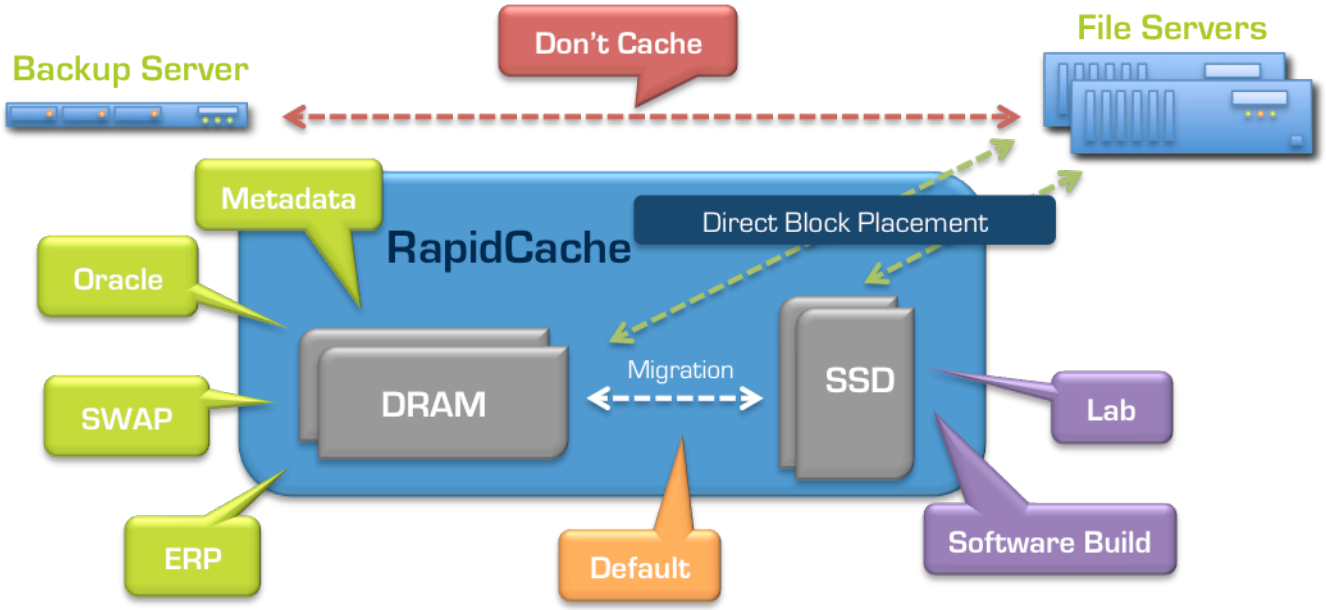


Intelligent Caching

The RapidCache solution provides the most advanced and effective means to allocate high-performance resources where and when they are required most. With its powerful packet inspection engine, IQ OS learns the content of data streams, and at wire-speed, makes in-flight decisions based on default or user-defined policies. Because there is no initial need to move

data to its assigned virtual tier, blocks are moved less frequently, which increases overall efficiency. However, as data demands change, IQ OS also considers frequency of use to promote or demote blocks between tiers (or completely out of cache).

Exceptions to this are user-created policies that "pin" data to a particular tier or the entire cache. A pinned policy means other data cannot evict the pinned data—regardless of frequency of use. This is incredibly useful for data that may not be accessed often, but is mission-critical when needed. In busy environments that do not support pinning, important, but seldom used data will never be read from cache because soon after it is cached, the data is evicted before it is needed again. Pinned profiles address this unwanted turnover.



Types of Policies

Default Policies

Default policies are built-in. They assign all metadata to the highest tier (currently DRAM) and all other data to a secondary pool with equal weight. Frequency of use will dictate if data is to be migrated between tiers; and with no user-defined profiles enabled, the default policy will hold all data.

File Policies

File policies include filenames, file extensions, file size, file server, and file system ID (FSID) in any combination with optional exclusions.

Example: Cache all *.dbf files less that 2GB from file server 192.168.2.88 and exclude file201.dbf

Client Policies

Client policies use IP addresses or DNS names with optional exclusions.

EXAMPLE: Cache all clients in IP range: 192.168.2.0/24 and exclude 192.168.2.31

NOTE: Policies can overlap to fine-tune cache performance and add flexibility.

Modifiers

Quota

Quotas optionally impose a limit on the amount of cache a policy consumes and can be specified by size or percent of overall cache. This can be particularly useful in multitenant storage environments to prevent one group from over-consuming resources.

Schedule

Define when a policy is to be activated or disabled based on a time schedule.

EXAMPLE: Activate profile “Nightly Software Build” at 9pm and disable at 6am.

Pin

Override normal block migration between tiers in the cache based on frequency of use and force data to stay within a particular tier or overall cache.

EAXAMPLE1: Pin profile “Database files” to DRAM – Files in this profile will not be migrated to a lower tier and more frequently used files will possibly be evicted sooner

EXAMPLE2: Pin Profile “VMware VMDK files” to Cache – Files in this profile are guaranteed to stay in cache, but will move between tiers based on usage.

Don’t Cache

Designate data not to be cached by file name of client request. This option is useful when dealing with data that is only read once, not critical, or which may change often behind the RapidCache.

EXAMPLES: Multiprotocol volumes, regularly snap restored volumes, and backup servers

Priority

Manually dictate the relative importance of policies to ensure data is evicted in the proper order. This allows user-defined priorities to assign quality of service based on business needs. Don’t cache, low, high, and pin are the available user-defined priorities.

EXAMPLE: “Backup Profile” = Don’t Cache

“Lab Profile” = Low Priority

“Production Profile” = High Priority

“SWAP Profile” = Pin to DRAM (Note: Pinned profiles override priority)

Cache Behavior

Eviction

Eviction is computed on cache priority from lowest to highest (don’t cache, default, low, high, and pin) and will start with the lowest and move to higher priority data only when the tier is full. Eviction from cache is driven first by priority and then usage. The lowest priority with the least accessed blocks will be evicted from cache first and the highest priority most used blocks will be evicted last.

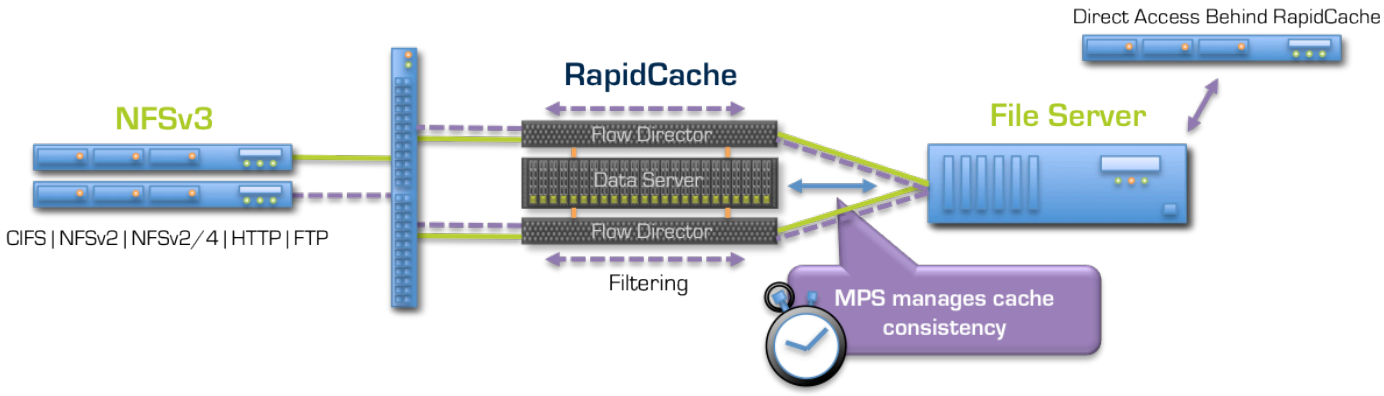
Migration

Migration of data within the cache is driven strictly by usage, and the most active data, without regard to priority, will migrate to the fastest cache tier. Likewise, as other data becomes more active, stale data will be demoted. Data pinned to a specified tier is excluded from migration.

Multi-Path Support (MPS)

Multi-Path Support is a mechanism for validating the data in RapidCache in the case where data behind RapidCache gets changed. MPS checks backend file server attributes at a configurable, predefined interval. Data may change when snap-restoring, using multiprotocol volumes (i.e. CIFS, NFSv2/4), or if there are clients directly modifying data on the backend file server. When a client reads a file, MPS evaluates its cache lease to determine whether it needs to check file server attributes. If not expired, the read will be served immediately from cache. If expired, MPS checks the backend file server to confirm no changes have occurred. If changes are found, MPS will pull the data from the file server, send it to the client, reset its lease, and update the cache.

With regular activity, file leases should never expire since they are updated on every NFS operation. Expiration only occurs on idle files. MPS timeout can be configured from 3 seconds to 24 hours.



High Availability

RapidCache can be configured with active/active redundant Data Servers and Flow Directors that keep synchronized through IQ OS with active heartbeats. If there is a node failure, the IQ OS removes the node from the cluster then instructs remaining nodes to rebalance and redistribute file responsibilities. If there is a Flow Director failure, traffic will continue through the redundant Flow Director.

Further high availability features have been incorporated in all RapidCache components to address all possible single points of failure.

Data Server

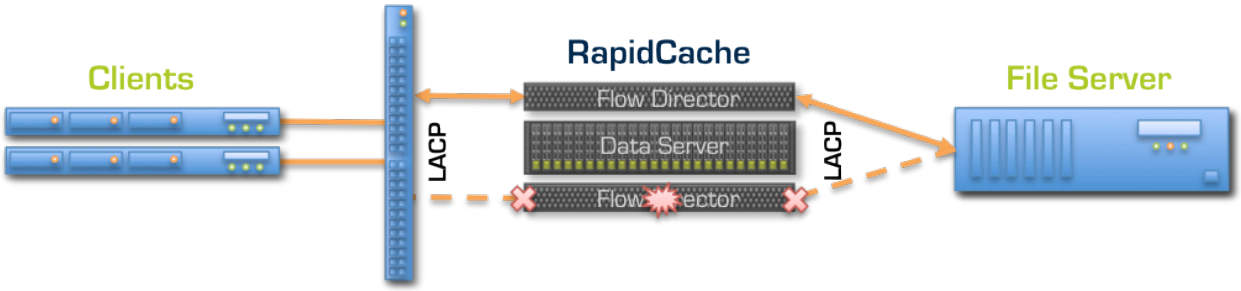
Each Data Server employs hot swappable power supplies, redundant fans, ECC memory and enterprise-level SSDs. Also, because no RAID is used, SSD failures only affect that portion of data and performance will not degrade for rebuilds—only capacity is reduced. SSDs are both hot replaceable and hot added.

Flow Director

Active/active redundant Flow Directors keep network traffic moving with hot swappable power supplies, redundant fans, and active heartbeats to Data Servers. In the unlikely case that a Flow Director becomes unresponsive, an internal hardware watchdog disables client/file server ports in order to facilitate failover on connected devices. The downed Flow Director is then forced to reload and can rejoin the cluster if found healthy.

Failover

Flow Directors do not initiate failover between redundant units because of their active/active configuration. Since it is the Data Servers that keep connection state between clients and file servers, network traffic can flow indiscriminately through either Flow Director—they are essentially equal bidirectional pathways to the same destination. Failover is negotiated between the client switch and individual file servers. IQ OS is transparent and facilitates failover with Link State Propagation and link aggregation protocols like LACP.



The Flow Director's internal hardware watchdog disables all network ports if it becomes unstable–allowing LACP to rebalance through alternate Flow Director

Fail to Wire

If all Data Server nodes fail to maintain heartbeat communications, the Flow Director automatically fails to wire and bypasses the nodes to send traffic directly to the backend file servers. Connections to backend file servers are never lost and client applications continue to run. Bypass mode can also be manually enabled for maintenance on Data Servers to provide uninterrupted service.

Alerts

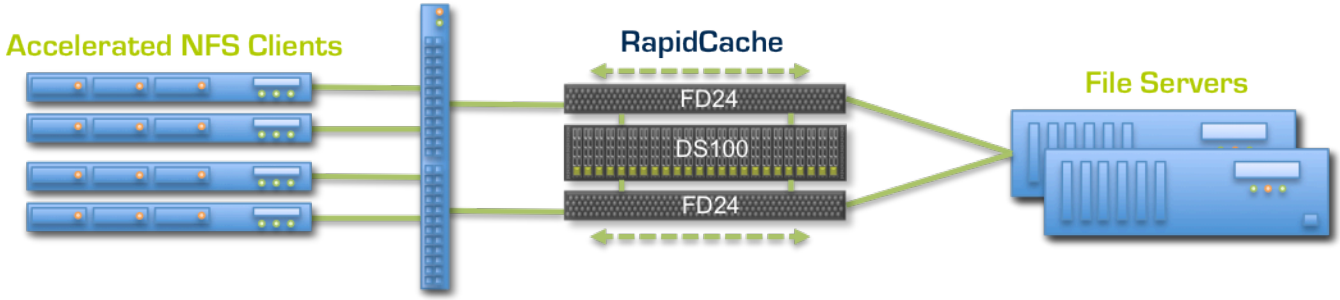
System health and events are reported through syslog or email and can be configured through RapidView. Five levels of message verbosity and multiple user accounts enable administrators the ability to fine-tune alerts. SNMP support is expected in future releases.

RapidView

RapidView does not have to be active for regular cache operation. It is only necessary when viewing statistics or modifying configurations. However, when it is connected, it can report environmental conditions across the cluster such as DIMM temperatures, fan failures, voltages, link status, and SSD failures.

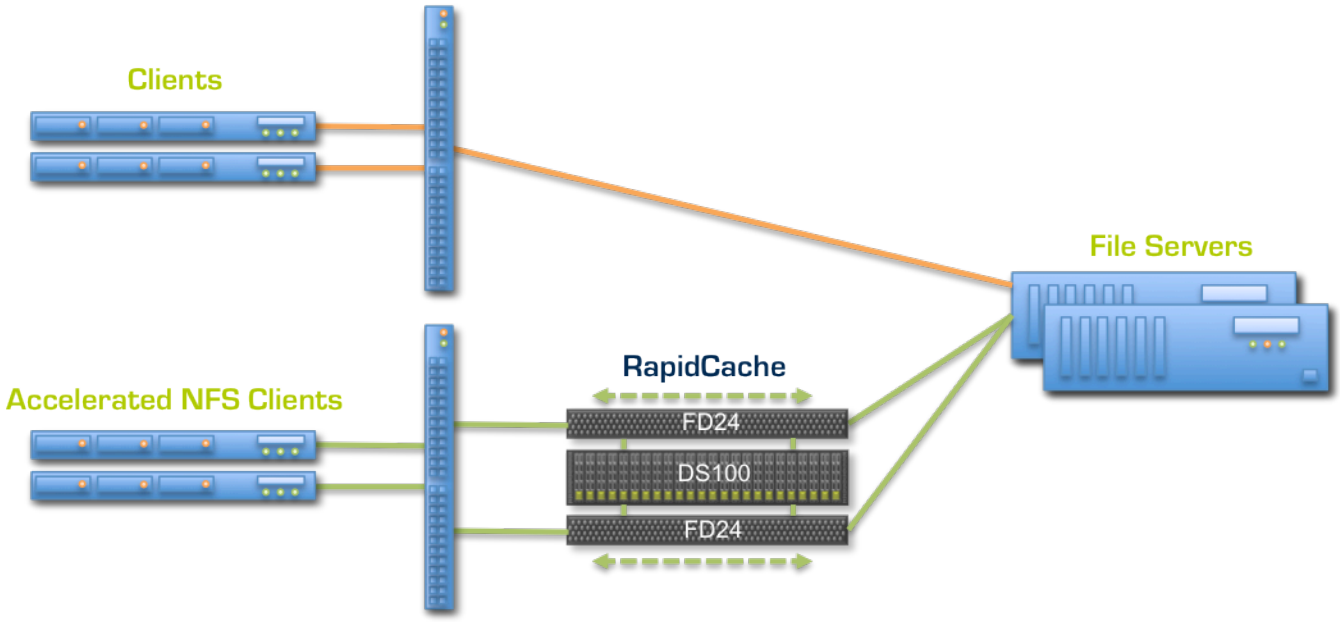
Deployment Options

Consolidated Data Path



This installation method processes all NFS traffic while filtering all other traffic to the file servers. RapidCache becomes the single physical network path, which provides the most caching flexibility through polices and the most complete analytics since it can review all traffic.

Multi Data Path



RapidCache can also be installed parallel to other network paths for specific departments, groups, clients or applications. By physically segregating accelerated clients through RapidCache, admins can choose specific clients that leverage the cache and analytics, leaving others untouched. With MPS, data remains consistent even if edited through different paths.

Conclusion

Cache IQ's RapidCache offers a scalable, easy to install solution that dynamically allocates performance to applications that need it most. Its unique packet inspection engine is the foundation that provides transparency, real-time prioritization of data access, and real-time statistics to prove the performance increases. Whether you are running applications like VMware over NFS or Oracle over NFS, the powerful RapidCache platform makes incorporating newer and faster media a seamless integration while offering the most flexibility and control.

CacheIQ's RapidView, simply takes the guess work out of managing network storage. With real-time views and historical data, RapidView provides both the insight into what is driving the storage network right now but also provides the historical detail required for the next high performance NAS purchasing decision.

Taken together, RapidCache and RapidView provide the IT manager the tools needed to tame network storage.

Cache IQ

phone 512-354-1254 info@cacheiq.com www.cacheiq.com