Caching in Shared Environments

Andy Walls

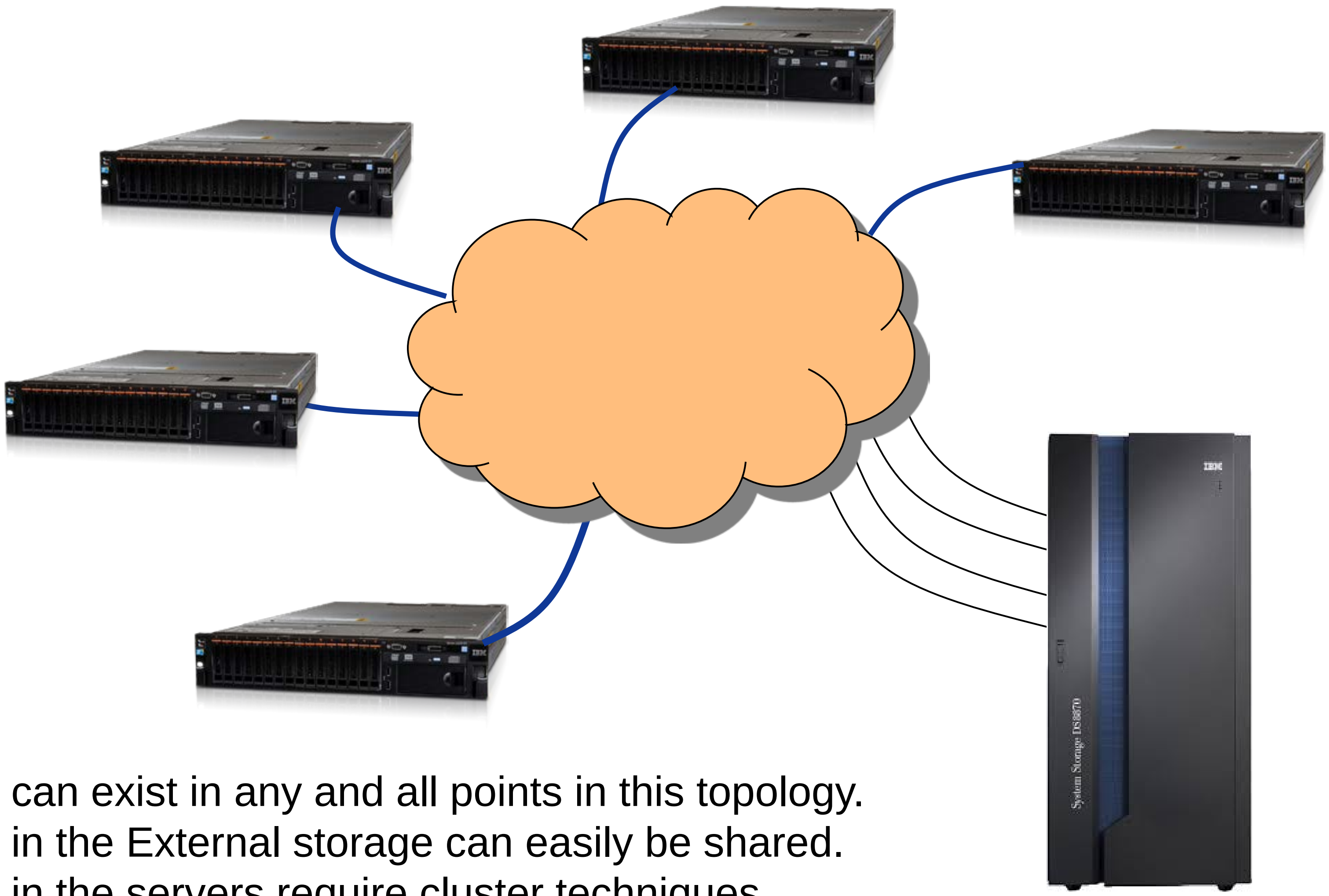
Distinguished Engineer, IBM Systems and Technology Group

CTO, Flash Systems and Technology

Flash Usage Topologies

- Flash Direct Attach Storage (DAS)
  - Temp data or non persistent storage
  - In a clustered Topology where robustness achieved by the application
- Flash in External Storage
  - All Flash Arrays
  - Hybrid flash using Automatic tiering or Caching
- Flash in servers and External storage
  - Tiering or caching with coherency managed by the servers
  - Tiering or caching with coherency managed by the storage

A Server/Storage Fabric



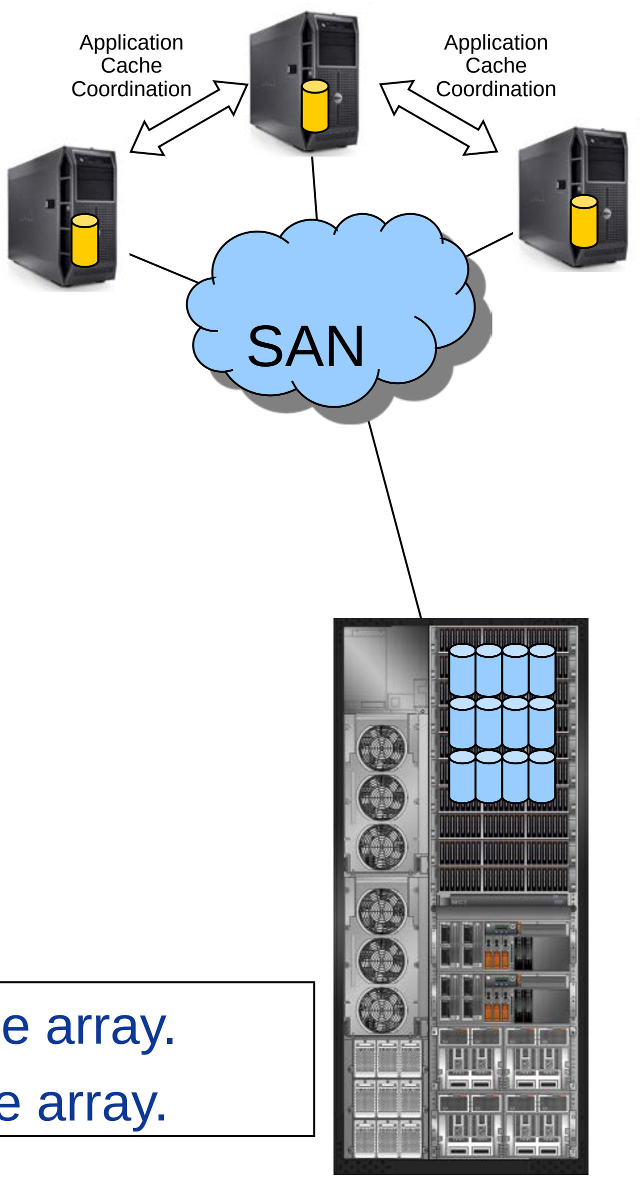
Flash can exist in any and all points in this topology.
Flash in the External storage can easily be shared.
Flash in the servers require cluster techniques
External storage is a natural point of coherency and control

Persistent Storage must have the following provided *Somehow*

- High Availability including concurrent code load, maintenance, fast write, redundancy, etc
- Snapshots
- Disaster recovery via remote copy.
- Security
- Simplified management
- Other features often required or desired like Compression, Dedupe, Cloning, low latency, etc.

Server based caching

- Off the shelf SSD's or Flash Cards installed in servers connected to the SAN.
- Data is copied, 'check-pointed' onto local storage volumes in each server.
- Application level clustering is required to keep all servers in sync.
- Data may be read only, or copied back to SAN periodically
- Advantages/Disadvantages
  - Lower initial/incremental flash cost, very low response times. Scale-out clustering.
  - Complex HA models


Application Cache Coordination
Application Cache Coordination
SAN

- Less accessed data is stored on the array.
- SSD may or may not be used in the array.

Tiering and Caching in External Storage

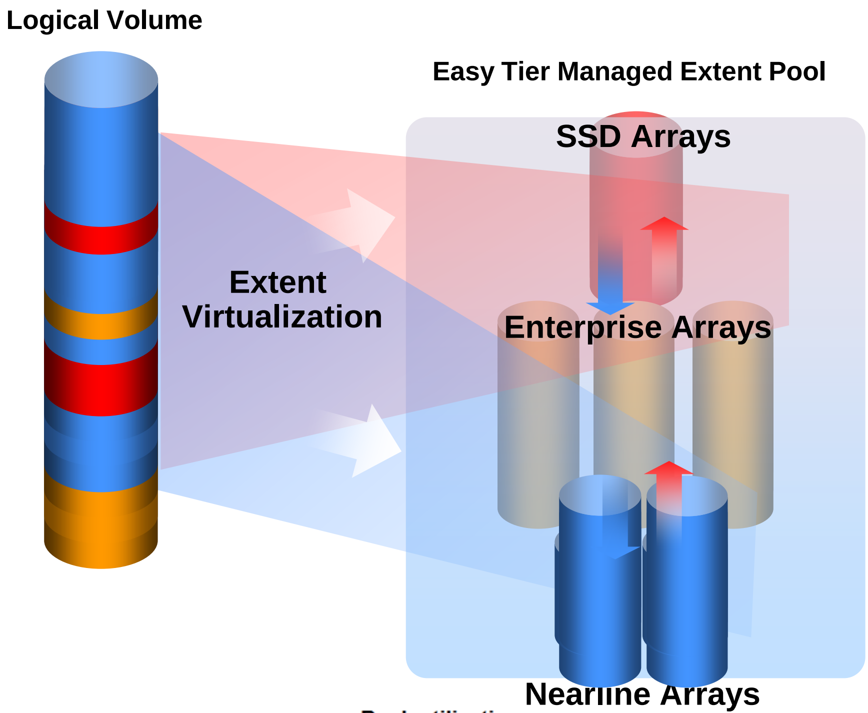
- External storage provides many advantages
  - Rich set of storage features.
  - Single point of coherency
  - Simplified management

Tiering and Caching has matured greatly offering a rich set of functionality.

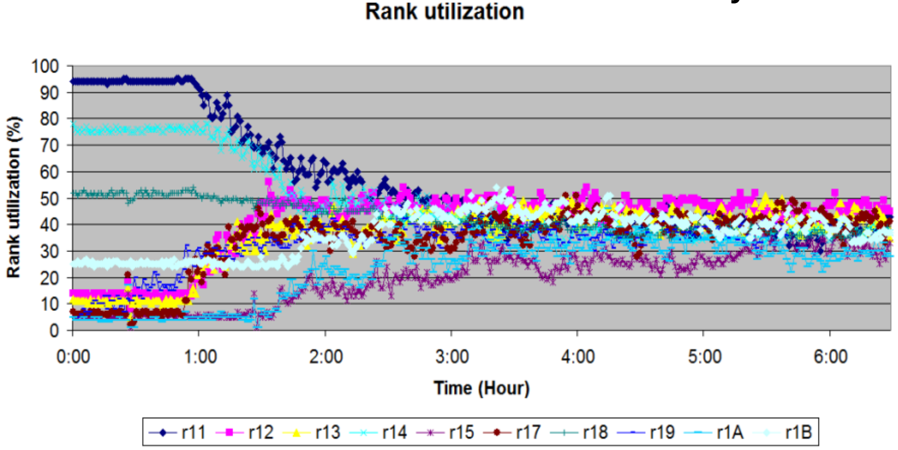
- Multiple tiers of Storage including different types of HDDs and SSDs
- Automated cross-tier performance or storage economics management for
  hybrid pools with multiple tools
- Automated *intra-tier performance management in both hybrid and homogenous pools* (auto-rebalance)
- Encryption support across a heterogeneous mix of devices
- Do no harm Migration
- Heat Map Transfer to Remote copy relationships
- Application guided tiering or caching.
- Support for thick or thinly provisioned volumes
- Read only caches with various population and destage algorithms
- Write back caches

Automated Tiering Illustration

- Full 3-tier technology support
  - Faster performance for hot data with SSDs
  - Cost savings (reduced footprint and $/GB) for cold data
  - Sophisticated algorithms automatically optimize each pool continuously



- Intra tier rebalancing and hot spot mitigation is important sid benefit
  - Supports both mixed-tier pools and single-tier pools

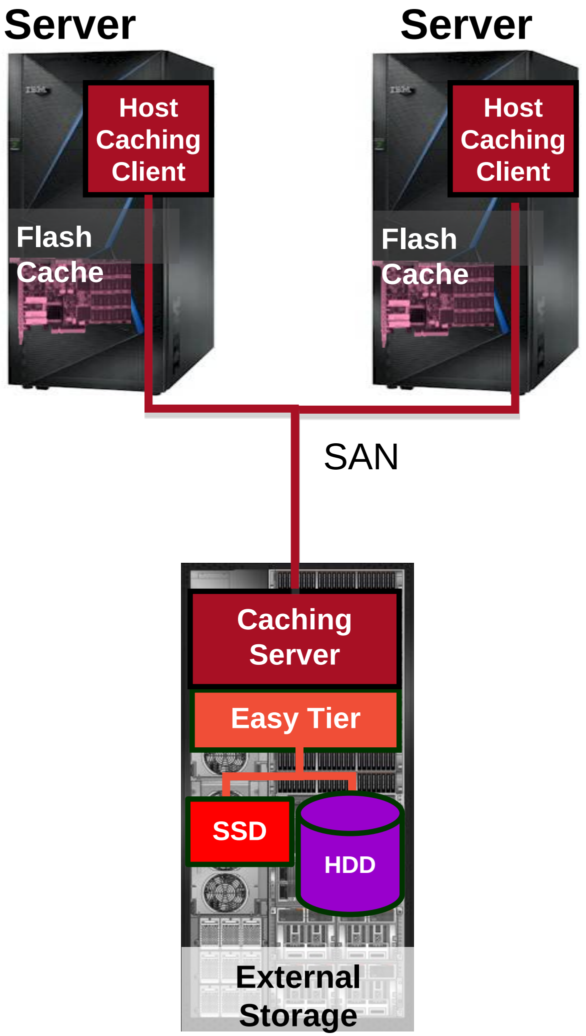


Collaborative Caching between Storage and Servers

- IBM Power p / DS8000 integration

- Easy Tier Server is an architecture:
  - A Host flash-cache based on IBM research technology using IBM flash hardware (Easy Tier Caching Client)
  - Easy Tier based algorithms for automatic application aware system-wide optimization (Easy Tier / Easy Tier Caching Server)
  - A proprietary protocol provides in band (SAN) communication and coherency

- Hosts (Caching Clients)
  - Work independently to cache their applications IO streams providing real-time performance enhancement
  - Cooperative Caching protocols allow system-aware caching to interface with Storage running the Cooperative Caching Server



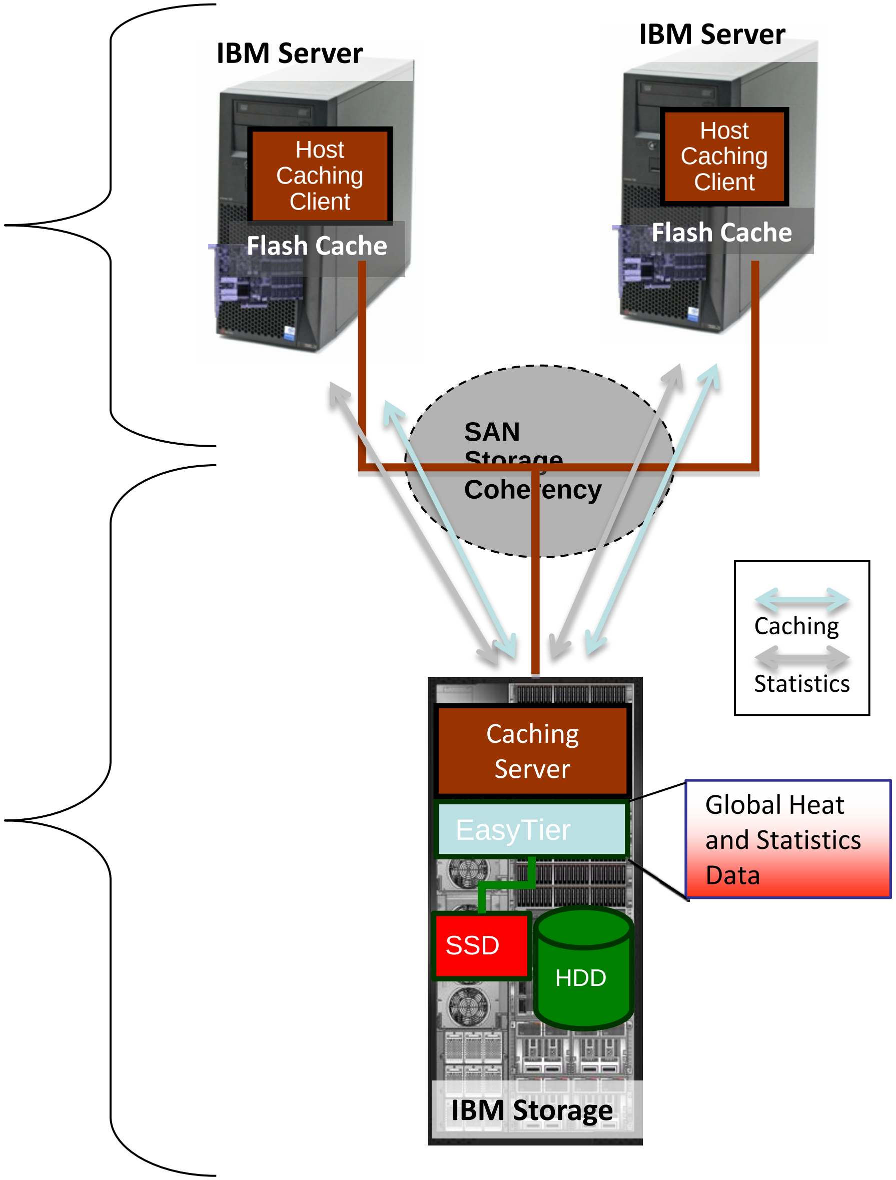
In this model caching in the host is a client managed by the external storage caching server.

Cache Client

- The driver in the Host SCSI stack determines what to cache based on access patterns
- The cache makes per-io decisions of what to keep in cache and what to evict
- Selection algorithms achieve high performance by remembering the latest evicted tracks when selecting what to promote into the cache

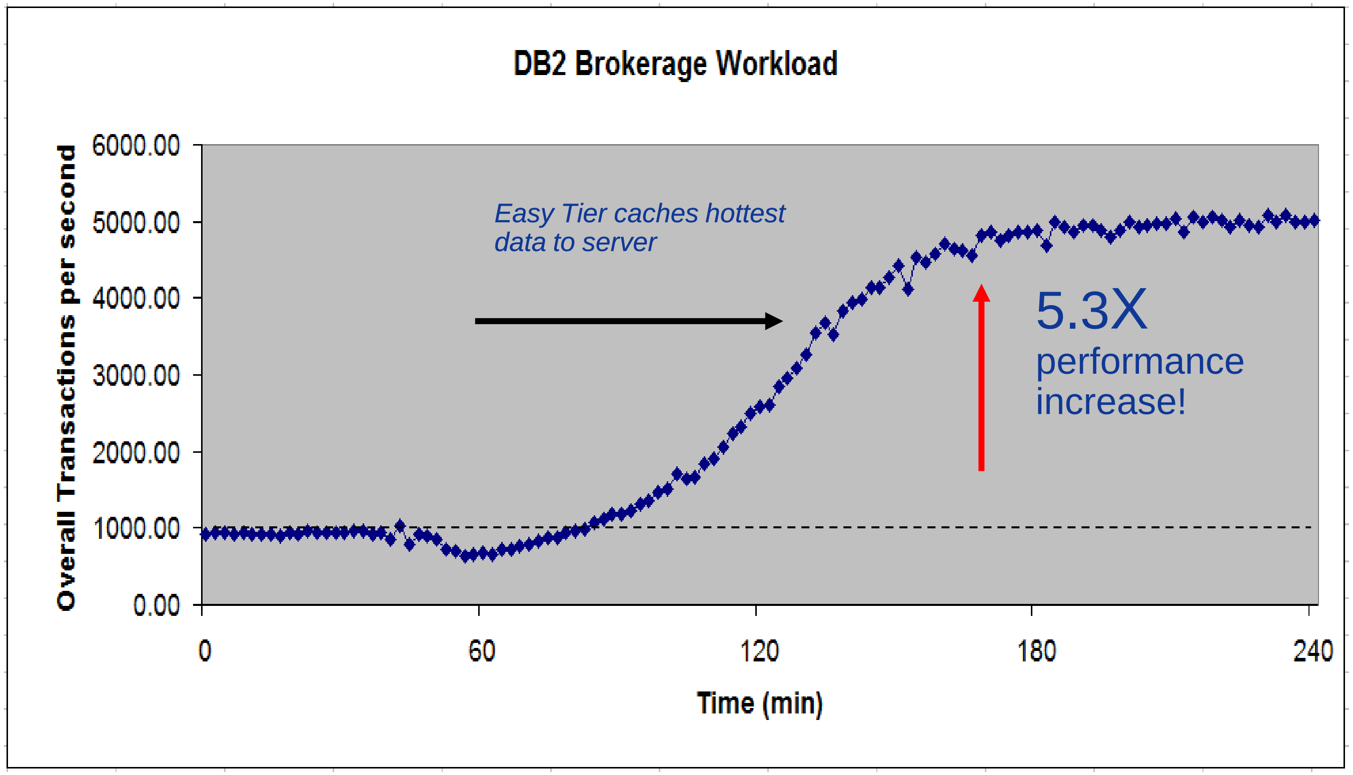
Cache Server

- While the Client Host SCSI stack driver does its own caching, it also receives 'advice' through Caching server specific communication
- The 'advice' is based on a unified view of the SAN storage, hosts, and their access patterns. It is a priority sorted list of extents
- The Client combines this advice with its own population list, resulting in both short and longer term cache population, higher hit ratio, and better storage solution optimization, including application aware storage.
- SAN based multi-host applications are enabled through the coherency protocol managed by the External Storage
- With coherency managed, storage advanced replication functions such as FlashCopy and remote mirroring are transparently enabled



Example of performance improvement

Up to 5X performance increase for DB2 banking brokerage workload



Base configuration is all-HDD with cache in Server not activated
IBM Power 770 server running AIX with 1 Ultra SSD I/O Drawer
DS8870 146GB 15K drives (RAID 5) with 2 1.3TB database volumes

Potential workloads for SAN Caching (roadmap view)

- **Performance acceleration for real time analytics workload:**
  - Very high Read/Write ratio
  - Very latency sensitive (such as Identity checking)
  - Highly parallel processing
- **Performance acceleration for large content data**
  - Big data is too big to fit in DRAM cache
  - Performance is improved by placing overflow data to local cache
- **Performance acceleration for OLTP workload**
  - Master copy is stored in primary SAN storage
  - Analytics apps load the data to local flash cache and operate on local cache copy.
- **VM consolidation and acceleration**
  - Enable more VM consolidation vie faster IO capability
  - VM acceleration by improve IO efficiency
- **Big Data**
  - Optimized for fast local primary copy access
  - Optimized for storage efficiency with network storage efficiency technology